

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 2002, 02:16:28 ; Search time 90 Seconds
(without alignments)
1041.683 Million cell updates/sec

Title: US-09-944-896-50

Perfect score: 2529
Sequence: 1 MLHPTSPGCHLAVLAL.....RNRICQFQHEHISRMQGS 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1154	45.6	334	6	Q9BE36 macaca fasc
2	397.5	15.7	156	4	Q9EH32 homo sapien
3	274.5	10.9	489	11	Q9J356 mus musculu
4	274.5	10.9	489	11	Q9ET66 mus musculu
5	272	10.8	415	4	Q8TCB8 mus musculu
6	257.5	10.2	497	4	Q9H0B8 mus musculu
7	257	10.2	371	4	Q96K61 mus musculu
8	254.5	10.1	255	11	Q9CW61 mus musculu
9	254.5	10.1	258	4	Q43692 mus musculu
10	254	10.0	523	13	Q98ST5 mus musculu
11	253	10.0	253	4	Q9H3Y0 mus musculu
12	252.5	10.0	258	13	Q98ST6 mus musculu
13	252	10.0	258	11	Q99MT7 mus musculu
14	245	9.7	233	4	Q96L06 mus musculu
15	240.5	9.5	434	11	Q9DTR3 mus musculu
16	240	9.5	188	11	Q9Z0U6 mus musculu

17	239	9.5	203	4	Q9H108 mus musculu
18	235.5	9.3	500	4	Q9H336 mus musculu
19	234	9.3	507	11	Q99MM6 mus musculu
20	231	9.1	266	4	Q969K2 mus musculu
21	228.5	9.0	301	5	Q950F6 mus musculu
22	226.5	9.0	1290	13	Q9W6E1 mus musculu
23	225	8.9	236	11	Q9DAG6 mus musculu
24	225	8.9	332	11	Q9CQ35 mus musculu
25	220	8.7	217	6	Q77719 mus musculu
26	219	8.7	415	5	Q44228 mus musculu
27	208	8.2	243	11	Q88205 mus musculu
28	207.5	8.2	243	11	Q9R1L4 mus musculu
29	207	8.2	244	11	Q91XA3 mus musculu
30	205	8.1	1321	4	Q14594 mus musculu
31	203	8.0	663	5	Q44247 mus musculu
32	201.5	8.0	250	11	Q9D259 mus musculu
33	196.5	7.8	380	5	Q9WYF2 mus musculu
34	196.5	7.8	392	5	Q960R5 mus musculu
35	195.5	7.7	315	4	Q9UPK6 mus musculu
36	191.5	7.6	249	6	Q9XSD3 mus musculu
37	189.5	7.5	1456	11	Q61830 mus musculu
38	189	7.5	207	5	Q20608 mus musculu
39	185.5	7.3	212	5	Q20609 mus musculu
40	185	7.3	168	10	Q40397 mus musculu
41	184.5	7.3	567	4	Q8WUL3 mus musculu
42	184.5	7.3	1140	4	Q96KG7 mus musculu
43	184	7.2	162	10	Q96344 mus musculu
44	183	7.2	199	13	Q8WU25 mus musculu
45	181.5	7.2	177	10	Q65157 mus musculu

ALIGNMENTS

RESULT 1
Q9BE36 PRELIMINARY; PRT; 334 AA.
ID Q9BE36
AC Q9BE36;
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 35.9 kDa protein.
OS Macaca fascicularis (Craab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxId=9341;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=CEREBELLUM CORTEX;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.,
RT "Isolation of full-length cDNA clones from macaque brain cDNA
libraries."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB060195; BAB4141.1; -;
DR HSSP: P04284; 1CFE.
DR InterPro: IPR001283; Allrgn_V5/TPx1.
DR Pfam: PF00188; SCP; 1.
DR PRINTS: PR00837; VSTPXLKE.
DR ProDom: PD000542; Allrgn_V5/TPx1; 1.
DR SMART: SM00198; SCP; 1.
KW Hypothetical protein.
SQ SEQUENCE 334 AA; 35917 MW; 37B37E3D29AC0D19 CRC64;

Query Match 45.6%; Score 1154; DB 6; Length 334;
Best Local Similarity 94.2%; Pred. No. 3.6e-102;
Matches 212; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLHPTSPGCHLAVLALGTTWAWPPOLOEAPMAGALNRKSFLLSLNRRLS 60
DB 1 MLHPTSPGCHLAVLAVLGLTTWAWPPOLOEAPMARALTRKESFLLSLNRRLS 60

```

QY 61 WVQPADMRRLDMSDSLQAQLAQAARALCGIPTPSLASSGLMRTLQVGMNQMLPAGLASF 120
    |||||
DB 61 WVQPPAAMRRRLDMSDSLQAQLAQAARALCGIPTPSLASSAPMHTLQVGMNQMLPAGSASF 120
QY 121 VEVVSLWFAEGORYSHAAEGECARNATCTHYTOLVWMTSSQLGGRHLCSAGQTAIEAFVC 180
    |||||
DB 121 VEVVSLWFAEGORYSHAAEGECARNATCTHYTOLVWMTSSQLGGRHLCSADQAAIEAFVC 180
QY 181 AYSFGNMWVNGKTIIPYKKGAMCSICTASVSGCFKAMDAGAGLC 225
    |||||
DB 181 AYSFGNMWVNGKTIIPYKKGAMCSICTASVSGCFKAMDAGAGLC 225

RESULT 2
Q96HH2 PRELIMINARY: PRT: 156 AA.
AC 096HH2:
DB 01-DEC-2001 (TREMBLrel. 19, Created)
DB 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DB 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DB Unknown (Protein for IMAGE:4178394) (Fragment).
DB Homo sapiens (Human).
DB Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DB NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008616; AA08616.1; -
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF000059; Lectin_C.1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
DR FT NON_TER
SQ SEQUENCE 156 AA: 17134 MW; 8A0F06A955C97306 CRC64;

Query Match 15.7%; Score 397.5; DB 4; Length 156;
Best Local Similarity 86.9%; Pred. No. 4e-30;
Matches 73; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

QY 372 GLYTKAKDSFRMATGEQATPSFPGPDHGLWLSAMGCVCELOASAFNNDDQ 431
    |||||
DB 82 GLYTKAKDSFRMATGEQATPSFPGPDHGLWLSAMGCVCELOASAFNNDDQ 132
    |||||
QY 432 RCKTRNRITCOFAOEHSRMWPGS 455
    |||||
DB 133 RCKTRNRITCOFAOEHSRMWPGS 156
    |||||

RESULT 3
Q9JJS6 PRELIMINARY: PRT: 489 AA.
AC 09JJS6:
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DB 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DB 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DB Cysteine-rich protease inhibitor.
DB 1200009H1LRK OR CRIP1.
DB Mus musculus (Mouse).
DB Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
DB NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Jang J.S., Hahn Y., Chung J.H.;
RA Identification of novel mouse cyteine-rich protease inhibitor
RT gene."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB046537; BAB03398.1; -
DR HSSP: P04284; 1CFE.
DR MGD; MG1:1921366; 1200009H1LRK.

```

```

DR InterPro: IPR001283; Allrgn_V5/7px1.
DR Pfam: PF00188; SCP.1.
DR PRINTS: PR00837; V5TPXLIKE.
DR PRODOM: PD000542; Allrgn_V5/7px1.1.
DR SMART: SM00198; SCP.1.
DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
DR KW Protease.
SQ SEQUENCE 489 AA: 52676 MW; 01C207BE12E3CB9D CRC64;

Query Match 10.9%; Score 274.5; DB 11; Length 489;
Best Local Similarity 32.3%; Pred. No. 1.1e-17;
Matches 75; Conservative 22; Mismatches 80; Indels 55; Gaps 9;

QY 13 LLAVLLALGTTWAEVWPPOLOEQAPMAGALNRKESFLLSLHNRLSWQPPADMRRL 72
    |||||
DB 7 LLALLLIATGPT-----TALTDEKQTMVDLHNOYRAQVSPASDMLQW 50
QY 73 DMSDSLQAQLAQAARALCGIPTPSLASSGLMRTLQVGMNQMLPAGLASFV-----EV 123
    |||||
DB 51 RMDDELAFAFAKAYAKOC-----VW-----GHNKERGRGENLFAITDEGMDVPLA 95
QY 124 VSLWFAEGORYSHAAEGECARNATCTHYTOLVWMTSSQLGGRHLCSAGQTAIEA----FV 179
    |||||
DB 96 VGNWHEHEHYNYNSTATCDPNOMCGHYTYQVWWSKTERIGCGSHCETLQGYEANIHLV 155
QY 180 CAYSPGNMWVNGKTIIPYKKGAMCSICTASVSGCFKAMDAGAGLCFPRMP 231
    |||||
DB 156 CAYEPGN--VKRK--PYOEGTPCSQCPGLAYS-C-----ENSLCEPRMP 196

RESULT 4
Q9ET66 PRELIMINARY: PRT: 489 AA.
AC 09ET66:
DB 01-MAR-2001 (TREMBLrel. 16, Created)
DB 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DB 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DB Cysteine-rich protease inhibitor.
DB 1200009H1LRK OR CRIP1.
DB Mus musculus (Mouse).
DB Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DB Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
DB NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Jang J.S., Hahn Y., Chung J.H.;
RA Genomic structure of murine cyteine-rich protease inhibitor gene."
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB046539; BAB03453.1; -
DR EMBL: AB046538; BAB03453.1; JOINED.
DR HSSP: P04284; 1CFE.
DR InterPro: IPR001283; Allrgn_V5/7px1.
DR Pfam: PF00188; SCP.1.
DR PRINTS: PR00837; V5TPXLIKE.
DR PRODOM: PD000542; Allrgn_V5/7px1.1.
DR SMART: SM00198; SCP.1.
DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
DR KW Protease.
SQ SEQUENCE 489 AA: 52664 MW; 01C2009712E3C76D CRC64;

Query Match 10.9%; Score 274.5; DB 11; Length 489;
Best Local Similarity 32.3%; Pred. No. 1.1e-17;
Matches 75; Conservative 22; Mismatches 80; Indels 55; Gaps 9;

QY 13 LLAVLLALGTTWAEVWPPOLOEQAPMAGALNRKESFLLSLHNRLSWQPPADMRRL 72
    |||||
DB 7 LLALLLIATGPT-----TALTDEKQTMVDLHNOYRAQVSPASDMLQW 50
QY 73 DMSDSLQAQLAQAARALCGIPTPSLASSGLMRTLQVGMNQMLPAGLASFV-----EV 123
    |||||
DB 51 RMDDELAFAFAKAYAKOC-----VW-----GHNKERGRGENLFAITDEGMDVPLA 95

```

Query	124	VSLMEFAGGGRSHAAAGCARNATCTHTYLTQVMAFPSSQLGGGRHLCSAGQTAIFA-----PV	179
Db	96	VGNMHEHEHYNTSTATCDPNQMGHTQVWVSKTERIGCGSHFCETLQGVETANHLHV	155
Qy	180	CAYSFGGNMEVNGKTIIPYKKGAMCSICTSVSGCFAMWADAGLCEVPNNP	231
Db	156	CNVEPPGN--VKGRK--PYQGTGTCSCDPLGYS-C-----ENSLCEPMNRP	196
RESULT 5			
08TCB8	08TCB8	PRELIMINARY;	PRT; 415 AA.
AC	08TCB8;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	HYPOTHETICAL 45.2 KDa protein (Fragment).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.		
RC	STRAUSBERG LUNG.		
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC022399; AAH22399.1; -		
KW	Hypothetical protein.		
FT	NON TER 1		
SO	SEQUENCE 415 AA; 45230 MW; 94506969C1607D38E3 CRC64;		
Query Match			
	Best Local Similarity	10.8%; Score 272; DB 4; Length 415;	
	Matches 74; Conservative 28; Mismatches 78; Indels 52; Gaps		
Qy	13	LLAVLLALGTTMAEWMPPQLEQAPMAGLNKKESEFLLSLNHRLSWQPPADMRRL	72
Db	18	LLPRLLLLVAT-----GPV-GALTDEKRLMELNMLYRAOVSPTASDMLHM	64
Qy	73	DMSDSLQALQARALGIPYPSLASGLMPTLQVGMNQLLPAGLASPV-----EV	123
Db	65	RWDELLAFAPAAIYAROC-----VW-----GHNKEGRGRENLFATIDEGMDVPLA	109
Qy	124	VSLMEFAGGGRSHAAAGCARNATCTHTYLTQVMAFPSSQLGGGRHLCSAGQTAIFA	179
Db	110	MEEMHHEHEHYNLSAATGSPQMGHTQVWVWATERIGCGSHFCETLQGVETANIELV	165
Qy	180	CAYSFGGNMEVNGKTIIPYKKGAMCSICTSVSGCFAMWADAGLCEVPNNP	231
Db	170	CNVEPPGN--VKGRK--PYQGTGTCSCDPLGYS-C-----GYHCKNSLCEPIGSP	210
RESULT 6			
09H0B8	09H0B8	PRELIMINARY;	PRT; 497 AA.
AC	09H0B8;		
DT	01-MAR-2001 (TREMBLrel. 16, Created)		
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	HYPOTHETICAL 55.9 KDa protein.		
GN	DKFZPA34B044.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.		
RC	TISSUE-TESTIS;		
RA	Wambuit R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.,		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL136861; CAB66795.1; -		
DR	HSSP; P04284; ICFE.		

```

DR InterPro: IPR001283; Allrgn_V5/TpX1.
DR InterPro: IPR004043; LCCL_dom.
DR Pfam: PF00188; SCP. 1.
DR PRINTS: PR00837; V5TPX1IKE.
DR ProDom: PD000542; Allrgn_V5/TpX1; 1.
DR SMART: SM00198; SCP. 1.
DR PROSITE: PS01010; SCP_Ag5_PRL_SC7_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 497 AA; 55919 MW; 1EADDCAB37C44ED CRC64;

Query Match 10.2%; Score 257.5; DB 4; Length 497;
Best Local Similarity 25.8%; Pred. No. 4,8e-16;
Matches 98; Conservative 45; Mismatches 126; Indels 111; Gaps 20.

OY 42 AINRKESELLSLNHRLSWVOPPADMRRLDMSDSLQAQRAALC---GIPPSLA 97
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 50 AIPREDKEELTMLHNKRLGVOVPOASNMMEYMTWDELEKSAAMAWSQCIWEHG-PT-SLL 107
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 98 SGLMTLTQVGMWMLLPAGLAFSEVYSLIMEEGQRYSHA-AGEC-----ARNATCTH 149
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 108 VSIGONLGAHMRYSRPG-----FHVOSWYDEVDYITPYRSECPWPCPERCGPMCTH 161
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 150 YTLQWLTAMSSQLCGSRHLG-----SAGQPIAEAFCAVSPGGMNVEVNGKTIIPKKGA 202
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 162 YIQIYWMATTNKGICGAVNCRKMTVMGEVEMNAV-YFVCVYSPKGNW--IGEA- -PYKNR 216
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 203 WCSLCTASVSG-----CFKANDH-----AGGLCEY-----PRNCP 232
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 217 PCSECPSPSGSGSCRNNLCYREFTYIPKPEPTDMENEYETAPIEDENHWMLQPRYMRPTPKR 276
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 233 RMSQCNH-----GRLNISCTCH--CPPGYTGKVCQVROSLOCVHGFRPEECS 279
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 277 KTSAAVNYMTQVVRCDTKMKDKCKSGTCNRYQCPAGCLNHKAKIFGTL-----FESSSS 330
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 280 -CVCIGVY-----GAQCATV--HFPHTCDLRI-DGD 309
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 331 ICRATHVGIILDDKGGVDTIRNGKVFPEVFKERHVOGLSLKRPSSPMVSKVYKQDID 390
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 310 CPWVSEADTYRYRARKCQR 329
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 391 CYTVAQLCPFEKRPATHCPR 410
   1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
O96K61 PRELIMINARY; PRT; 371 AA.
AC O96K61;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE CDNA FLJ14489 f15b, clone MAMMA1002881, weakly similar to glioma
   pathogenesis-related protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Iisgai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuna S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuko Y., Sasaki N.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK021395; BAB55081.1; -
DR InterPro: IPR001283; Allrgn_V5/TpX1.
DR InterPro: IPR004043; LCCL_dom.
DR Pfam: PF00188; SCP. 1.
DR ProDom: PD000542; Allrgn_V5/TpX1; 1.
DR PROSITE: PS01010; SCP_Ag5_PRL_SC7_2; UNKNOWN.1.
SQ SEQUENCE 371 AA; 42207 MW; 8D23FBE14F53E85C CRC64;

```

```

Query Match 10.2%; Score 257; DB 4; Length 371;
Best Local Similarity 28.8%; Pred. No. 3,6e-16;
Matches 89; Conservative 36; Mismatches 102; Indels 82; Gaps 17;

QY 42 ALNRKESFLLSLNRLRSWOPPADMRRLDWSLSLAQLAQAALC---GIPTPSIA 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 50 AIPREDKEEIIILHNKKLKGVOPOQASNNMEYMTWDELEKSAAMASQCIWEGH-PT-SLL 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 98 SGLWLTGVGNWQMLPPGLASFEVSVLMPFEGGRYSHA-AGEC-----ARNATCTH 149
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 VSGNGLAHMGRYSPG-----FHVOSWIDEVDYTYPIPSSECNPMPCPERSGPMCTH 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 150 YTVLWVATTSOLGCGRHLC-----SAGQTAIEAFVCAVSPGNGMEVNGKTIIPYKKA 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 162 YTVLWVATTNKIGCAVNTCRKMTVGEVWENAV-YFVCNYSPPKGMW--IGEA- PYKNGR 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 203 WCSISCTASVSG-----CFKAMH-----AGLGCV-----PRNDC 232
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 217 PCSPEPPSYGSGSRNNLCYREETYPKPEFDEMNVEAPRIEENHWLAPRVMPRTPK 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 223 RMSCONH-----GRLNISTGCH-CPPGYTGRYCOVRSCLGCHGRFREECSS 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 KTSANVMTQVVRDPTKKKDRCKSGSTCNRYCCPAGCLMHNKKIPTLT-----FTSSSS 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 280 -CVCDDICYG 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 ICRRAIHYG 339

RESULT 8
09CWC1 PRELIMINARY; PRT; 255 AA.
AC 09CWC1;
DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE 2410114014Rik protein (RIKEN CDNA 2410114014 gene).
GN GN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
   [1]
RP
RN
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kodota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojtuna N., Carninci P., de Bonaldo M.F.,
RA Rastbach S., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmberg L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:585-590(2001).
RN
RP
SEQUENCE FROM N.A.
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK010768; BAB27168.1; -
DR EMBL: BC025083; AAH25083.1; -
DR HSSP: P04284; ICPE.

```

```

DR      MGI:1920940; 241011401.R1K.
DR      InterPro: IPR001283; Allrgn_V5/Tpx1.
DR      Pfam: PF00188; SCP, 1.
DR      PRINTS: PR00837; V5TPXLIKE.
DR      ProDom: PD00542; Allrgn_V5/Tpx1, 1.
DR      SMART: SM00198; SCP, 1.
DR      PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
SQ      SEQUENCE 255 AA; 29128 MW; A836ECBBACBD20A CRC64;

Query Match          10.1%; Score 254.5; DB 11; Length 255;
Best Local Similarity 31.0%; Pred. No. 3.9e-16;
Matches 72; Conservative 30; Mismatches 81; Indels 49; Gaps 11.

QY      14 LAVLLALIGTTWAELVSPQLOEQADMGALN-----RKESFL-LISLHNR 59
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      1 MOVILLAVI-----VW-----MASSSSSFTASTLPDITNEDFIKECVYHNQ 45

QY      60 SWVQPPADMRRLKSDSLAQLAQAARALCGIP-TPSLASGLMRTL-QVGMNMLIPAC 117
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      46 SKVSPAPRNMYLMSWDPKLQIAKAWTKSCERKHPDLSRTHPNFTALGENIWLGS 105

QY      118 ASFEVLSLTFEGGGRYSHAAGECARNNKTCFHYQLVMAVTSQQLCGRLTCAAGQTAL 177
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      106 FSVSALSAIAMEELIHYDFSTRKCH--VCGHYTQVWADSVKYLCAVQLCPNGAN 159

QY      178 FVCAASPGCNGEVNNGKTIIPYKGAWCSSICTASVGCFCFAMDHAGLCVEPR 229
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      160 FICDYGPGNPT-----WPKQGTGTCSDCPED-DKCLNS-----LCINPR 199

```

RESULT	ID	PRELIMINARY:	PRT:	258 AA.
9	043692			
10	043692			
11	01-JUN-1998	(TrEMBLrel. 06, Created)		
12	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)		
13	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)		
14	25 kDa	trypsin inhibitor.		
15	De	Homo sapiens (Human).		
16	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
17	OC	Mammalia; Euthera; Primates; Carnivora; Hominiidae; Homo.		
18	OX	NCBI_TaxID=9606;		
19	RP	SEQUENCE FROM N.A.		
20	RA	MEDLINE=98146272; PubMed=9473672;		
21	RA	Yamakawa T., Miyata S., Ogawa N., Koshikawa N., Yasumitsu H.,		
22	RA	Kanamori T., Miyazaki K.		
23	RT	"DNA cloning of a novel trypsin inhibitor with similarity to		
24	RT	pathogenesis-related proteins, and its frequent expression in human		
25	RT	brain cancer cells."		
26	RL	Biochim. Biophys. Acta 1395:202-208(1998).		
27	EMBL	D45027; BAA25066.1; -		
28	DR	HSSP; P04284; ICFE.		
29	DR	InterPro: IPR001283; Allrgn_V5/WpX1.		
30	DR	Pfam: PF00188; SCP. 1.		
31	DR	PRINTS: PRO0837; VSTPLIKE.		
32	DR	ProDom: PD000542; Allrgn_V5/WpX1. 1.		
33	DR	SMART: SM00198; SCP. 1.		
34	DR	PROSITE: PS01010; SCP_AGS_PRL_SCT_2. 1.		
35	DR	SEQUENCE 258 AA; 29065 MW; 191545831637795F CRC64;		
36	Query Match	10.1%; Score 254.5; DB 4; Length 258;		
37	Best Local Similarity	29.2%; Pred. No. 3.9e-16;		
38	Matches 70; Conservative	33; Mismatches 66; Indels 71; Gaps 11;		
39	43	LNKRESEFLLSLNHRLESWQPPADNRRLRSDSLAQDAARALCGIPPSLASGLW-101		
40	62	ISQNDMIDAIIDYHNOVGKVEPPAANNEYMWMDENLAKSADAWATC-----IMD 111		
41	102	-----RTLDYGNMMDLPAGLAFVEVYSLMAEQRYSNAGE-----CARNA---T 146		
42	112	HGSPYLRLFL--GNLSVFRGGRYSIIQLVKWYDEVKDYAFRRDQDNCPRMKSGRM 169		

Query Match 10.0%; Score 252.5; DB 13; Length 258;
 Best Local Similarity 28.8%; Pred. No. 6.1e-16;
 Matches 69; Conservative 34; Mismatches 66; Indels 71; Gaps 11;

QY 43 LNRKSEFLSLHNRRLSVQPPADMRRLDMSLSLAQLAQAARALGIPPSLASGLW-101
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC014603; AAI14603.1;
 DR InterPro: IPR001283; Allrgn_V5/Tpx1.
 DR InterPro: IPR001760; Opsln.
 DR Pfam; PF00188; SCP; 1.
 DR Prodom; PD000542; Allrgn_V5/Tpx1; 1.
 DR PROSITE; PS01009; SCP_AGS_PRL_SCT_1; UNKNOWN_1.
 DR PROSITE; PS01010; SCP_AGS_PRL_SCT_2; UNKNOWN_1.
 SQ SEQUENCE 233 AA; 26118 MW; C6BDD4B8DB5B91D2 CRC64;

RESULT 13

QY 43 LNRKSEFLSLHNRRLSVQPPADMRRLDMSLSLAQLAQAARALGIPPSLASGLW-101
 ID 099MM7 PRELIMINARY; PRT; 258 AA.
 AC 099MM7;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Sugarcrisp.
 GN P115.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-21184121; PubMed-11287197;
 RA Smith D.M., Collins-Racie L.A., Marigo V.A., Roberts D.J., Davis N.M.,
 RA Hartman C., Schweltzer R., Lavallie E.R., Gomer L., McCoy J.,
 RA Tabin C.J.;
 RT "Cloning and expression of a novel cysteine-rich secreted protein
 RT family member expressed in thyroid and pancreatic mesoderm within the
 RT chicken embryo."
 RL Meach. Dev. 102:223-226(2001).
 DR EMBL; AF329196; AAK16494.1;
 DR HSSP; P04284; ICFE.
 DR MGD; MGI:1934659; P115.
 DR InterPro: IPR001283; Allrgn_V5/Tpx1.
 DR Pfam; PF00188; SCP; 1.
 DR PRINTS; PR00837; VSTPXLKE.
 DR PRODOM; PD000542; Allrgn_V5/Tpx1; 1.
 DR SMART; SM00198; SCP; 1.
 DR PROSITE; PS01010; SCP_AGS_PRL_SCT_2; 1.
 SQ SEQUENCE 258 AA; 29150 MW; 2FA94A9A2562640 CRC64;

Query Match 10.0%; Score 252; DB 11; Length 258;
 Best Local Similarity 32.2%; Pred. No. 6.8e-16;
 Matches 65; Conservative 31; Mismatches 60; Indels 46; Gaps 10;

QY 43 LNRKSEFLSLHNRRLSVQPPADMRRLDMSLSLAQLAQAARALGIPPSLASGLW-101
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
 QY 200 KGAMCSLCTASVSG-----CF 215

DB 225 VGVPCSSCPSPYVGACTDNLCT 246

RESULT 14

QY 43 LNRKSEFLSLHNRRLSVQPPADMRRLDMSLSLAQLAQAARALGIPPSLASGLW-101
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC014603; AAI14603.1;
 DR InterPro: IPR001283; Allrgn_V5/Tpx1.
 DR InterPro: IPR001760; Opsln.
 DR Pfam; PF00188; SCP; 1.
 DR Prodom; PD000542; Allrgn_V5/Tpx1; 1.
 DR PROSITE; PS01009; SCP_AGS_PRL_SCT_1; UNKNOWN_1.
 DR PROSITE; PS01010; SCP_AGS_PRL_SCT_2; UNKNOWN_1.
 SQ SEQUENCE 233 AA; 26118 MW; C6BDD4B8DB5B91D2 CRC64;

Query Match 9.7%; Score 245; DB 4; Length 233;
 Best Local Similarity 34.2%; Pred. No. 2.8e-15;
 Matches 63; Conservative 24; Mismatches 75; Indels 22; Gaps 8;

QY 52 LSHNRRLSVQPPADMRRLDMSLSLAQLAQAARALGIPPSLASGLWRTL---QVG 107
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
 QY 166 HLC-SAGQTAIEAFVCAVSPGMEVNGKTIIPYKKGWCSLCTASVSGCFKAMDAGL 224
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
 QY 154 AMCPNLGASALFVCMVGNPAGNF-----ANMPYVREGSCSLCSNE-EKCYK-----NL 202
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
 QY 225 CEVP 228
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
 QY 203 CKNP 206

RESULT 15

QY 43 LNRKSEFLSLHNRRLSVQPPADMRRLDMSLSLAQLAQAARALGIPPSLASGLW-101
 ID 096L06 PRELIMINARY; PRT; 233 AA.
 AC 096L06;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE Similar to RIKEN CDNA 1700011E04 gene.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
 QY 203 CKNP 206


```

XX 20-NOV-1998; 98US-0109304.
PR 05-JAN-1998; 98US-0070440.
PR 29-APR-1998; 98US-0083500.
PR 22-MAY-1998; 98US-0086414.
PR 10-JUN-1998; 98US-0088742.
PR 10-NOV-1998; 98US-0107783.
XX
XX (GETH ) GENENTECH INC.
XX
XX Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;
XX Roy MA, Wood WI;
XX WPI; 1999-430385/36.
XX N-PSDB; AAX87260.
XX
XX Antibody against proteins expressed in neoplastic cells, useful for
XX tumor diagnosis and treatment
XX
XX Example 1; Fig 14; 162pp; English.
XX
XX This sequence represents human PRO347 (UHQ306), a 50.5 kDa protein
XX (PI 8.44) encoded by the novel cDNA clone DNA44176 (see AAX87260).
XX Amplification of DNA44176 was observed in various tumour lines,
XX suggesting a role in tumour formation and growth. Antagonists
XX (e.g. antibodies) directed to PRO347 may have use in cancer therapy.
XX The invention identifies 14 genes (see AAX87254-67) that are amplified
XX in the genome of tumour cells. Such amplification is expected to be
XX associated with overexpression of the gene product and to contribute
XX to tumorigenesis. The encoded proteins (see AAX06477-90) may be
XX useful targets for the diagnosis and/or treatment (including
XX prevention) of certain cancers, and may act as predictors of the
XX prognosis of tumour treatment. Antibodies that bind the proteins
XX are claimed and used in claimed cancer diagnostic kits.
XX
XX Sequence 455 AA:
XX
XX Query Match 100.0%; Score 2529; DB 20; Length 455;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-183;
XX Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MHPEPSRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLS 60
XX 1 MHPEPSRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLS 60
XX 61 WQPPAADRRRLDWSLSLAQARAALGIPPSLASGLMRTLQVGMNQLPAGIASF 120
XX 61 WQPPAADRRRLDWSLSLAQARAALGIPPSLASGLMRTLQVGMNQLPAGIASF 120
XX 121 VEVVSLMFEAGQRTSHAEGECARNATCTHTYTLVWATSSOLGCGRHLCASGQTAIEAFVC 180
XX 121 VEVVSLMFEAGQRTSHAEGECARNATCTHTYTLVWATSSOLGCGRHLCASGQTAIEAFVC 180
XX 121 VEVVSLMFEAGQRTSHAEGECARNATCTHTYTLVWATSSOLGCGRHLCASGQTAIEAFVC 180
XX 121 VEVVSLMFEAGQRTSHAEGECARNATCTHTYTLVWATSSOLGCGRHLCASGQTAIEAFVC 180
XX 181 AVSPGNNVWNGKTTIPYKKGAMCSICTASVSGCFKAMDAGGICEVPRNCRSCQNHG 240
XX 181 AVSPGNNVWNGKTTIPYKKGAMCSICTASVSGCFKAMDAGGICEVPRNCRSCQNHG 240
XX 181 AVSPGNNVWNGKTTIPYKKGAMCSICTASVSGCFKAMDAGGICEVPRNCRSCQNHG 240
XX 241 RINISTCHCHCPGTYGRYCOVRCSLQCVHGRFREECSCVCDIGYGGAOCATKVPHEPFH 300
XX 241 RINISTCHCHCPGTYGRYCOVRCSLQCVHGRFREECSCVCDIGYGGAOCATKVPHEPFH 300
XX 241 RINISTCHCHCPGTYGRYCOVRCSLQCVHGRFREECSCVCDIGYGGAOCATKVPHEPFH 300
XX 301 TCDLRIDGCFWVSSADYTYRRARKCQKRGVLAQIKSOKVODILAFYIGRETTNEVT 360
XX 301 TCDLRIDGCFWVSSADYTYRRARKCQKRGVLAQIKSOKVODILAFYIGRETTNEVT 360
XX 301 TCDLRIDGCFWVSSADYTYRRARKCQKRGVLAQIKSOKVODILAFYIGRETTNEVT 360
XX 301 TCDLRIDGCFWVSSADYTYRRARKCQKRGVLAQIKSOKVODILAFYIGRETTNEVT 360
XX 361 DSDFEIRNFWIGLITYTAKDSFRWANGEGHAFTSFAGQDPNHLVLSAMGNCVEL 420
XX 361 DSDFEIRNFWIGLITYTAKDSFRWANGEGHAFTSFAGQDPNHLVLSAMGNCVEL 420
XX 361 DSDFEIRNFWIGLITYTAKDSFRWANGEGHAFTSFAGQDPNHLVLSAMGNCVEL 420
XX 421 QASAAFNMDORCKTRNRYTCQFAQEHISRWGPGS 455
XX 421 QASAAFNMDORCKTRNRYTCQFAQEHISRWGPGS 455
XX

```

```

XX RESULT 2
XX AAY17828
XX ID AAY17828 standard; Protein; 455 AA.
XX
XX AC AAY17828;
XX
XX 12-AUG-1999 (first entry)
XX
XX Human PRO347 protein sequence.
XX
XX Human: PRO protein; tumour necrosis factor family; TNF; cytokine;
XX secreted protein; transmembrane protein; inflammation disorder.
XX
XX Homo sapiens.
XX
XX WO9928462-A2.
XX
XX 10-JUN-1999.
XX
XX 01-DEC-1998; 98WO-US25108.
XX
XX 25-FEB-1998; 98US-0075945.
XX 03-DEC-1997; 97US-0067411.
XX 11-DEC-1997; 97US-0069278.
XX 11-DEC-1997; 97US-0069334.
XX 11-DEC-1997; 97US-0069335.
XX 12-DEC-1997; 97US-0069425.
XX 16-DEC-1997; 97US-0069694.
XX 16-DEC-1997; 97US-0069696.
XX 16-DEC-1997; 97US-0069702.
XX 17-DEC-1997; 97US-0069870.
XX 17-DEC-1997; 97US-0069873.
XX 18-DEC-1997; 97US-0068017.
XX 05-JAN-1998; 98US-0070440.
XX 09-FEB-1998; 98US-0074086.
XX 09-FEB-1998; 98US-0074092.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Chen J, Goddard A, Gurney AL, Wood WI;
XX Yuan J;
XX WPI; 1999-371118/31.
XX N-PSDB; AAX80053.
XX
XX Nucleic acids encoding PRO secreted and transmembrane proteins
XX
XX Claim 12; Fig 23; 123pp; English.
XX
XX The present invention describes nucleic acids encoding PRO secreted and
XX transmembrane proteins used therapeutically. The PRO proteins have
XX cytosolic, anti-inflammatory, anti-proliferative and immunosuppressive
XX activity. The proteins and polynucleotides can be used in therapy,
XX identification of homologues, raising antibodies and design of proteins
XX and primers. They can be used in a range of diseases related to proteins
XX that they have homology with, e.g. a PRO protein having homology to
XX complement proteins may be used in inflammatory responses.
XX
XX Sequence 455 AA:
XX
XX Query Match 100.0%; Score 2529; DB 20; Length 455;
XX Best Local Similarity 100.0%; Pred. No. 1.7e-183;
XX Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 MHPEPSRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLS 60
XX 1 MHPEPSRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLS 60
XX 61 WQPPAADRRRLDWSLSLAQARAALGIPPSLASGLMRTLQVGMNQLPAGIASF 120
XX 61 WQPPAADRRRLDWSLSLAQARAALGIPPSLASGLMRTLQVGMNQLPAGIASF 120
XX

```


421 QASAFNNDQRCRTNRNYICQFAOEHSRWGPGS 455
 421 QASAFNNDQRCRTNRNYICQFAOEHSRWGPGS 455

RESULT 4
 AA41266
 ID AA41266 standard; Protein: 446 AA.

AA41266;
 31-JAN-2000 (first entry)

Human T139 protein.

Human: T139 polypeptide; immune system disorder; spermatogenesis;
 sperm-egg fusion; testicular disorder; testicular cancer; gene mapping.

Homo sapiens.

MO9954343-A2.

28-OCT-1999.

23-APR-1999; 99WO-US08896.

23-APR-1998; 98US-0065661.

(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.

Holtzman D;

WPI: 1999-633969/54.

DR N-PSDB; AA232399; AA233300.

Human T139 nucleic acids and polypeptides, useful for treating
 proliferative disorders associated with aberrant T139 expression or
 activity

Claim 9; Fig 1; 115pp; English.

This represents a human T139 polypeptide. The T139 polypeptide can be
 deposited with ATCC under accession number 98694. The T139 polypeptides
 and polynucleotides can be administered therapeutically or
 prophylactically to treat/prevent disorders associated with aberrant T139
 expression or activity, especially proliferative or differentiative
 disorders, e.g. of the immune system. They can be used to modulate
 spermatogenesis, e.g. as a contraceptive to decrease spermatogenesis or
 be useful to treat testicular disorders e.g. testicular cancer. They may also
 be useful to treat testicular disorders e.g. testicular cancer. The
 polypeptides may be used to identify selectively binding compounds which
 may be useful for detecting the polypeptides in samples, and identifying
 compounds modulating polypeptide activity. The polynucleotides are useful
 for producing probes or primers that selectively hybridize to the
 polynucleotides which may be useful for detecting the polynucleotides in
 a sample, gene mapping, identifying cells or tissues expressing aberrant
 T139 levels; determining if a gene has been mutated or deleted to
 identify subjects at risk for or having a disorder associated with T139
 expression or activity and to monitor therapeutic interventions; and for
 producing antisense sequences for therapeutic administration to modulate
 T139 expression.

Sequence 446 AA;

Query Match 96.8%; Score 2447.5; DB 20; Length 446;

Best local Similarity 97.1%; Pred. No. 2,4e-177;

Matches 442; Conservative 1; Mismatches 3; Indels 9; Gaps 1;

1 MLHPETSPGRGHLAVLLALIGTVAEWPPOLQDAPAGALNKRESFLLSLHRLRS 60
 1 MLHPETSPGRGHLAVLLALIGTVAEWPPOLQDAPAGALNKRESFLLSLHRLRS 60

QY 61 WVPPEADMRRLWSDSLAQLAARALCGIPTPSLASCIMRLQVGMNQLPAGLASF 120
 Db 61 WVPPEADMRRLWSDSLAQLAARALCGIPTPSLASCIMRLQVGMNQLPAGLASF 120
 QY 121 VEVSILMFAEGGRYSNAGECANATCTHTYTLVWATSSQLGGRHCSAGOTAEAFVC 180
 Db 121 VEVSILMFAEGGRYSNAGECANATCTHTYTLVWATSSQLGGRHCSAGOTAEAFVC 180
 QY 181 AYSFGNMEVNTKTIIPYKGAWSCLTASVSGCFKAMDHAGICEVPRNCRMSQNHG 240
 Db 181 AYSFGNMEVNTKTIIPYKGAWSCLTASVSGCFKAMDHAGICEVPRNCRMSQNHG 240
 QY 241 RLNISTCHCHCPPEYGRICQVRCISLQCVHGRFREECSCVCDIGYGACATVHPEFH 300
 Db 241 RLNISTCHCHCPPEYGRICQVRCISLQCVHGRFREECSCVCDIGYGACATVHPEFH 300
 QY 301 TCDLRIDGCEFWSSPADYRRARMKCQRKGVLAQIKSQKVDILAFYLRLETTNEVT 360
 Db 301 TCDLRIDGCEFWSSPADYRRARMKCQRKGVLAQIKSQKVDILAFYLRLETTNEVT 360
 QY 361 DSDPETRNFWIGLTYRTAKDSFRWATGEHOAFTSFARQPDNGLWLSAAMGFCVEL 420
 Db 361 DSDPETRNFWIGLTYRTAKDSFRWATGEHOAFTSFARQPDNGLWLSAAMGFCVEL 420
 QY 421 QASAFNNDQRCRTNRNYICQFAOEHSRWGPGS 455
 Db 421 QASAFNNDQRCRTNRNYICQFAOEHSRWGPGS 455

RESULT 5
 AA06575
 ID AA06575 standard; Protein: 446 AA.

AA06575;

25-SEP-2001 (first entry)

Human protein having hydrophobic domain, HP10760.

Human: hydrophobic domain; gene therapy; nutritional supplement;
 cell proliferation; immunomodulatory; autoimmune disorder; antimicrobial;
 multiple sclerosis; rheumatoid arthritis; insulin-dependent diabetes;
 haemopoiesis; tissue growth activity; Parkinson's disease; cystostatic;
 Huntington's disease; Alzheimer's disease; chemotactic; chemokinetic;
 haemostatic; thrombolytic; tumour growth inhibitor; anabolic;
 contraceptive; antiinfertility; antiinflammatory.

Homo sapiens.

Key Location/Qualifiers

Peptide 1..27
 FT /label= Signal_peptide
 FT 28..446
 FT Protein
 FT /note= "Mature human protein with hydrophobic domain"

MO200149728-A2.

12-JUL-2001.

28-DEC-2000; 2000WO-JP09359.

06-JAN-2000; 2000JP-0000585.

06-JAN-2000; 2000JP-0000588.

11-JAN-2000; 2000JP-0002299.

03-FEB-2000; 2000JP-0026862.

03-MAR-2000; 2000JP-0058367.

(PROT-) PROTEGENE INC.
 PA (SAGA) SAGAMI CHEM RES CENT.
 PA Kato S, Kimura T;
 PI WPI: 2001-418355/44.

DR N-PSDB; AAD12570.

XX Human proteins with hydrophobic domains and the nucleic acids encoding
PT them, useful for preventing diagnosing and treating e.g. cancer,
XX Alzheimer's and inflammation -

PS Claim 1; Page 230-232; 563pp; English.

CC The present sequence is human protein with hydrophobic domain,
CC HPI0760. The polynucleotide and polypeptide of the invention
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The polynucleotides
CC into a host cell and culturing the polypeptide, by inserting the nucleic acids
CC polynucleotides and its complementary sequences may also be used as DNA
CC probes in diagnostic assays and also used in gene therapy. The
CC polypeptides may also be used as antigens in the production of antibodies
CC and in assays to identify modulators of polypeptide expression and
CC activity. The polypeptides and nucleic acids may be used as nutritional
CC supplements, to modulate cytokine and cell proliferation activity, to
CC modulate immune stimulation or suppression (e.g. for the treatment of
CC microbial infections and autoimmune disorders such as multiple sclerosis,
CC rheumatoid arthritis and insulin-dependent diabetes), to modulate
CC haematopoiesis, to modulate tissue growth activity (e.g. for the
CC treatment of Parkinson's disease, Huntington's disease and Alzheimer's
CC disease), to modulate actin and inhibit activity (e.g. for controlling
CC fertility), to modulate chemotactic and chemokinetic activity, to
CC modulate haemostatic and thrombolytic activity, to modulate receptor
CC ligand activity, to modulate inflammation and to inhibit tumour growth.

XX Sequence 446 AA;

Query Match Best Local Similarity 96.8%; Score 2447.5; DB 22; Length 446;

Matches 442; Conservative 1; Mismatches 3; Indels 9; Gaps 1;

DB 1 MLHPETSPGRGHLAVLLALIGTTWAEVWPOLQOAPMAGLNKRESFLLSLNRLRS 60
DB 1 MLHPETSPGRGHLAVLLALIGTTWAEVWPOLQOAPMAGLNKRESFLLSLNRLRS 60
QY 61 WVOPPADMRRLDMSDSLQAQARAALCGIPPSLASGLMRTLOVGNMOLLPLGLASF 120
DB 61 WVOPPADMRRLDMSDSLQAQARAALCGIPPSLASGLMRTLOVGNMOLLPLGLASF 120
QY 121 VEVVSLMFABGQRYSHAAEGCARNAATCTHYTQVWATSSQLCGRHLCSAQATAIEAFVC 180
DB 121 VEVVSLMFABGQRYSHAAEGCARNAATCTHYTQVWATSSQLCGRHLCSAQATAIEAFVC 180
QY 181 AYSFGNMWENVGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNCRMSCONHG 240
DB 181 AYSFGNMWENVGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNCRMSCONHG 240
QY 241 RLNISTCHCHCPGYTGRYQVRCSLQCVHGRFREECSCVCDIGYGAQCATVHPFFH 300
DB 241 RLNISTCHCHCPGYTGRYQVRCSLQCVHGRFREECSCVCDIGYGAQCATVHPFFH 300
QY 301 TCDLRIDGDCFMWSSSEADTYTTRAMKCORRGVLAQIKSQVODILAFYLGRLTETNEVT 360
DB 301 TCDLRIDGDCFMWSSSEADTYTTRAMKCORRGVLAQIKSQVODILAFYLGRLTETNEVT 360
QY 361 DSDPETRNFNIGLTYTAKOSFRWATGEOHAFSTPAFGOPDNHGLVWLSAAMGPNCEVL 420
DB 361 DSDPETRNFNIGLTYTAKOSFRWATGEOHAFSTPAFGOPDNHGLVWLSAAMGPNCEVL 420
QY 421 QASAAFMWNDQCKTRNNRYTCQFAOEHSIRWPGS 455
DB 421 QASAAFMWNDQCKTRNNRYTCQFAOEHSIRWPGS 455
QY 412 QASAAFMWNDQCKTRNNRYTCQFAOEHSIRWPGS 446
DB 412 QASAAFMWNDQCKTRNNRYTCQFAOEHSIRWPGS 446

RESULT 6
AAB64994
ID AAB64994 standard; protein; 446 AA.
XX

AC AAB64994;

XX 23-MAR-2001 (first entry)

XX Human secreted protein #2.

XX Secreted protein; gene therapy; vaccine; cancer; leukemia;
KW autoimmune disease; allergy; inflammation; graft rejection;
KM hyperproliferation; cardiovascular; infection.

OS Homo sapiens.

PN WO200075375-A1.

PD 14-DEC-2000.

PF 02-JUN-2000; 2000WO-US15187.

XX 07-JUN-1999; 99US-0137725.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;

PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;

WI; 2001-061741/07.

PT Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for
PT preventing, diagnosing and/or treating cancers and for promoting wound
PT healing -

PS Claim 1; Page 446-448; 530pp; English.

CC The present invention relates to 26 secreted human proteins. The
CC proteins may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate polypeptide expression.
CC For example, they may be used in gene therapy or in vaccines.
CC Typical of diseases which are potentially treatable are cancers
CC (including leukemia), autoimmune diseases, allergies, inflammation,
CC graft rejection, hyperproliferation, cardiovascular diseases
CC (particularly critical limb ischemia and coronary disease) and any
CC involving abnormal angiogenesis, neurodegeneration and/or
CC infectious diseases.

XX Sequence 446 AA;

Query Match Best Local Similarity 96.8%; Score 2447.5; DB 22; Length 446;

Matches 442; Conservative 1; Mismatches 3; Indels 9; Gaps 1;

QY 1 MLHPETSPGRGHLAVLLALIGTTWAEVWPOLQOAPMAGLNKRESFLLSLNRLRS 60
DB 1 MLHPETSPGRGHLAVLLALIGTTWAEVWPOLQOAPMAGLNKRESFLLSLNRLRS 60
QY 61 WVOPPADMRRLDMSDSLQAQARAALCGIPPSLASGLMRTLOVGNMOLLPLGLASF 120
DB 61 WVOPPADMRRLDMSDSLQAQARAALCGIPPSLASGLMRTLOVGNMOLLPLGLASF 120
QY 121 VEVVSLMFABGQRYSHAAEGCARNAATCTHYTQVWATSSQLCGRHLCSAQATAIEAFVC 180
DB 121 VEVVSLMFABGQRYSHAAEGCARNAATCTHYTQVWATSSQLCGRHLCSAQATAIEAFVC 180
QY 181 AYSFGNMWENVGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNCRMSCONHG 240
DB 181 AYSFGNMWENVGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNCRMSCONHG 240
QY 241 RLNISTCHCHCPGYTGRYQVRCSLQCVHGRFREECSCVCDIGYGAQCATVHPFFH 300
DB 241 RLNISTCHCHCPGYTGRYQVRCSLQCVHGRFREECSCVCDIGYGAQCATVHPFFH 300
QY 301 TCDLRIDGDCFMWSSSEADTYTTRAMKCORRGVLAQIKSQVODILAFYLGRLTETNEVT 360
DB 301 TCDLRIDGDCFMWSSSEADTYTTRAMKCORRGVLAQIKSQVODILAFYLGRLTETNEVT 360

[illegible]

Pt		Novel TANGO polypeptides and nucleic acid molecules useful as
Pt		modulating agents in regulating cellular processes and for diagnosing
Pt		and treating heart, liver, lung, kidney, inflammatory and cellular
Pt		proliferative disorders -
Xx		
Xx		Claim 49; Fig 1; 138pp; English.
Cc		The invention relates to nucleic acids encoding a variety of proteins
Cc		human T139 (TANGO-139), T125 (TANGO-125), T110 (TANGO-110), murine T175
Cc		(TANGO-175), human T175 or murine WDMN-2, having diagnostic, preventive,
Cc		therapeutic and other uses. Polypeptide of the invention has the ability
Cc		to inhibit a proteinase activity, to modulate cell-cell interactions,
Cc		haematopoiesis and the ability to modulate clotting. Polypeptide and
Cc		polynucleotide of the invention are useful for diagnosing and treating
Cc		disorder characterised by their aberrant expression or activity. The
Cc		antibodies are useful as modulating agents in regulating a variety of
Cc		cellular processes e.g. cell proliferation and/or cell differentiation.
Cc		TANGO-139 is useful for treating kidney defects such as kidney failure,
Cc		TANGO-125 is useful in wound healing and for treating cancer, TANGO-110
Cc		is useful for treating neoplasia, TANGO-177 or WDMN-2 is useful for
Cc		treating cancer, are useful to treat pancreatic disorders, such as
Cc		pancreatitis, cerebrovascular disease, and tumours, and injury or trauma
Cc		to the brain. TANGO-125, 110, 175 molecules treat heart disorders, e.g.,
Cc		ischemic heart disease, cardiovascular disorders, such as ischaemic
Cc		heart disease). TANGO-139, 125, 110 and 175 molecules are useful to treat
Cc		renal (kidney) disorders, such as glomerular disease (e.g., acute and
Cc		chronic glomerulonephritis), TANGO-175 is useful to treat uterine
Cc		diseases, hyperplasia of the endometrium. TANGO-110 is useful to treat
Cc		spleen, e.g., the fetal spleen, associated diseases and disorder. TANGO-
Cc		125 treats prostate disorders, such as inflammatory diseases, Crohn's
Cc		disease and tumours. TANGO-139, 125, 110, 175 or WDMN-2 are useful for
Cc		treating proliferative disorders, inflammatory disorders. TANGO-175, or
Cc		WDMN-2 activity also include apoptotic disorders, rheumatoid arthritis,
Cc		systemic lupus erythematosus, insulin-dependent diabetes mellitus,
Cc		immune-related disorders, e.g., immunodeficiency disorders, viral
Cc		disorders, cell growth disorders' e.g., cancers and inflammatory
Cc		disorders and apoptotic disorders. The nucleic acids of the invention
Cc		are used in gene therapy. The present sequence is human T139 protein.
Cc		
Xx		
Sq	Sequence	446 AA:
Qy	Query Match	96.8%; Score 2447.5; DB 23; Length 446;
Qy	Best Local Similarity	97.1%; Pred. No. 2.4e-177;
Matches	442; Conservative	1; Mismatches 3; Indels 9; Gaps 1;
Dy	1 MLHPTSPGSGRLILAVLLALLGTTFWEVMPPOLOEQAPMAGALNRKESFLLLSCHNRURS	60
Dy	1 MLHPETSGRGILLAVLLALLGTAMEVWPPOLOEQAPMAGALNRKESFLLLSLHNRLRS	60
Qy	61 WVOPRAADWRILDWSDSLQAOLAQRALACSLPPPSLASGLMRITLYQGNNMOLPLPGLASF	120
Dy	61 WVOPRAADRRLDWSDSLQAOLAQRALAIGIPPLSASGLMRITLYQGVNMOLLPLPGLASF	120
Qy	121 VEVVSIMWFEGORYSHAAGECARNACTHTLTQLMWATSSQLCGCSNLCSAGOTALAEAFVC	180
Dy	121 VEVSIMWFEGORYSHAAGECARNACTHTTLQYMAISSQLCGRHLCISAGAATALEAFVC	180
Qy	181 AYSPGGNMFVNCKTIIPYKKGANWCSTLTSASYSGCFKAMDHAGLCEVPBPNCRMSQNHG	240
Dy	181 AYSPGNMFVNCKTIIIPYKKGANWCSTLTSASVGCFCFKAMDHAGLCEVBPNCRMSQNHG	240
Qy	241 RLNIISTCHCHCPGYGRYCYOVRCSLOCVMHRFRFBEDSCVCDIGYGGAQCATKYHFPH	300
Dy	241 RLNISITCHEHCPCGYGRYCVRCSLOCVMHRFREBEESCVCDDIGYGGAQCATKYHFPH	300
Qy	301 TODDLRDGGCCFMWSSSEADITYRRARKMKCORRKGVAIAIKSQKODILAFYLGRLETNEYI	360
Dy	301 ITDDLRLDGDCFMWSSSEADITYRRARMKCORRKGVLAIAIKSQAYODILAFYLGILETTNEVI	360
Qy	361 DSDDEFTRNWIGILTLYKAANDSFRRWTGEHQATFSFAFGOPDNHGLVWISAAMFGNCVEL	420
Dy	361 DSDFETRNRWIGILTLYKAAKDSFRMATGEHQAFFTSPAFGOPDNH-----GFNGCNVEL	411

QY 421 QASAFNMNDORCKTRNRYICQFAOEHSRMGPGS 455
 DB 412 QASAFNMNDORCKTRNRYICQFAOEHSRMGPGS 446

RESULT 8
 ID AAM93326 standard; Protein: 446 AA.

XX AAM93326;
 AC AAM93326;
 DT 06-NOV-2001 (first entry)

XX Human polypeptide, SEQ ID NO: 2851.

XX Human, full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELF-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR N-PSDB; AAK94246.

PT 830 Primers useful for synthesizing full length cDNA clones and their

XX use in genetic manipulation -

XX Claim 8; SEQ ID NO 2851; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA

XX clones. 830 cDNA molecules encoding a human protein have been

XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

XX cDNA are useful for clarifying the function of the protein encoded by

XX the cDNA. The full length clones were obtained by construction of full

XX length enriched cDNA libraries that were synthesised by the oligo-capping

XX method. The primers enable the production of the full length cDNA easily

XX without any special methods. The present sequence is a polypeptide

XX encoded by a full length human cDNA of the invention.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 446 AA;

XX

DB

QY

DB

QY

QY 181 AYPGNGWVNGKTIIPKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNRCMSQNHG 240
 DB 181 AYPGNGWVNGKTIIPKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNRCMSQNHG 240
 QY 241 RLNIStCHCPCPGYTGRCVRCSLQCVHGRFREECSCVDIGYGAOCATKHFPPH 300
 DB 241 RLNIStCHCPCPGYTGRCVRCSLQCVHGRFREECSCVDIGYGAOCATKHFPPH 300
 QY 301 TCDLRIDGDCFMVSSADTYTARMKCORRKGVLAAQIKSQKVDILAFYLGRLFTTNEVT 360
 DB 301 TCDLRIDGDCFMVSSADTYTARMKCORRKGVLAAQIKSQKVDILAFYLGRLFTTNEVT 360
 QY 361 DSDPETRNFVIGLTYKTAKDSFRMATGSHOAFSTFAFGQPDNH-----GFGNCVEL 411
 DB 361 DSDPETRNFVIGLTYKTAKDSFRMATGSHOAFSTFAFGQPDNH-----GFGNCVEL 411
 QY 421 QASAFNMNDORCKTRNRYICQFAOEHSRMGPGS 455
 DB 412 QASAFNMNDORCKTRNRYICQFAOEHSRMGPGS 446

RESULT 9
 ID AAY41267 standard; Protein: 420 AA.
 XX AAY41267;
 AC AAY41267;
 DT 31-JAN-2000 (first entry)

XX Human T139 mature protein sequence.
 DE Human; T139 polypeptide; immune system disorder; spermatogenesis;
 KW sperm-egg fusion; testicular disorder; testicular cancer; gene mapping.
 XX Homo sapiens.
 OS WO9954343-A2.
 XX 28-OCT-1999.
 PD 23-APR-1999; 99WO-US08896.
 PF 23-APR-1998; 98US-0065661.
 PR 23-APR-1998; 98US-0065661.
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PI Holtzman D;
 DR WPI: 1999-633969/54.
 DR N-PSDB; AAZ23299, AAZ23300.
 XX Human T139 nucleic acids and polypeptides, useful for treating
 PT proliferative disorders associated with aberrant T139 expression or
 PI activity -
 PS Claim 2; Page 114-115; 115pp; English.

XX This is the mature protein sequence of the human T139 polypeptide. The
 CC T139 polypeptide can be expressed by standard recombinant methodology.
 CC The T139 cDNA insert is deposited with ATCC under accession number
 CC 98694. The T139 polypeptides and polynucleotides can be administered
 CC therapeutically or prophylactically to treat/prevent disorders associated
 CC with aberrant T139 expression or activity, especially proliferative or
 CC differentiative disorders, e.g. of the immune system. They can be used to
 CC modulate spermatogenesis, e.g. as a contraceptive to decrease
 CC spermatogenesis or to treat disorders related to defects in sperm-egg
 CC fusion. They may also be useful to treat testicular disorders e.g.
 CC testicular cancer. The polypeptides may be used to identify selectively
 CC binding compounds which may be useful for detecting the polypeptides in
 CC samples; and identifying compounds modulating polypeptide activity. The
 CC polynucleotides are useful for producing probes or primers that
 CC selectively hybridize to the polynucleotides which may be useful for
 CC detecting the polynucleotides in a sample, gene mapping; identifying

cells or tissues expressing aberrant T139 levels; determining if a gene has been mutated or deleted to identify subjects at risk for or having a disorder associated with T139 expression or activity and to monitor therapeutic interventions; and for producing antisense sequences for therapeutic administration to modulate/prevent T139 expression.

Sequence 420 AA:

Query Match 91.6%; Score 2315.5; DB 20; Length 420;
Best Local Similarity 97.2%; Pred. No. 2.3e-167;
Matches 417; Conservative 1; Mismatches 2; Indels 9; Gaps 1;

27 EWPMPOLQOAPMAGALNKRSEFLLSLHNRLRSWVOPPADMRRLDMSLSLAQAOARA 86
|||||
Db 1 EWPMPOLQOAPMAGALNKRSEFLLSLHNRLRSWVOPPADMRRLDMSLSLAQAOARA 60
|||||
Qy 87 ALGCIPTPSLASGLMRTLOVGMNMQLLPAGLASFEVYSLWFAEGORYSHAAGCARNAT 146
|||||
Db 61 ALGCIPTPSLASGLMRTLOVGMNMQLLPAGLASFEVYSLWFAEGORYSHAAGCARNAT 120
|||||
Qy 147 CTHYQVLWATSSQLCCGRRLCSAGQIAFAVAYSPGGWVEVNGKTIIPYKKAQMSL 206
|||||
Db 121 CTHYQVLWATSSQLCCGRRLCSAGQIAFAVAYSPGGWVEVNGKTIIPYKKAQMSL 180
|||||
Qy 207 CTASVSGCFKAMPDHAGGLCEVPRNPRMCSGHNRLNSTCHCHCPGTYGRCVCSL 266
|||||
Db 181 CTASVSGCFKAMPDHAGGLCEVPRNPRMCSGHNRLNSTCHCHCPGTYGRCVCSL 240
|||||
Qy 267 QCVHGRFREECSCVCDIGYGAQCATKVFPHFTCDLRIDGCEFMVSSEADTYRRARK 326
|||||
Db 241 QCVHGRFREECSCVCDIGYGAQCATKVFPHFTCDLRIDGCEFMVSSEADTYRRARK 300
|||||
Qy 327 CQRKGVLAKIKQKQVODILAFYLGLETTNEVYDSDPETNFTVGLTYKTKAKDSFRAT 386
|||||
Db 301 CQRKGVLAKIKQKQVODILAFYLGLETTNEVYDSDPETNFTVGLTYKTKAKDSFRAT 360
|||||
Qy 387 GEHOATSFAGQPDNNGVLWLSNAAGFNCVELOASAFNMNDORCTRNRYICQFAOE 446
|||||
Db 361 GEHOATSFAGQPDNNGVLWLSNAAGFNCVELOASAFNMNDORCTRNRYICQFAOE 411
|||||
Qy 447 HISMWPGS 455
|||||
Db 412 HISMWPGS 420
|||||
RESULT 10
AAE21078
ID AAE21078 standard; Protein; 420 AA.
XX AAE21078;
AC AAE21078;
XX
XX 01-JUL-2002 (first entry)
DT
DE Human T139 (TANGO-139) mature protein.
XX
KW Human: haematopoiesis; clotting; kidney failure; wound healing; cancer;
neoplasia; pancreatic disorder; pancreatitis; cerebrovascular disease;
heart disorder; ischaemic heart disease; neuroprotective; vulvar;
cardiovascular disease; ischaemic heart disease; immunosuppressive;
glomerular disease; glomerulonephritis; uterine disorder; hyperplasia;
fetal spleen; prostate disorder; inflammatory disease; Crohn's disease;
proliferative disorder; gynaecological; haemostatic; antibacterial;
systemic lupus erythematosus; immunodeficiency disorder; antiaesthetic;
cytotoxic; nephrotropic; antidiabetic; cerebroprotective; tranquilliser;
hypotensive; tumour; injury; trauma; antineoplastic; vasodilator;
apoptotic disorder; rheumatoid arthritis; candida; renal disorder;
hepatotropic; antiproliferative; antiallergic; dermatological; virucide;
T139.
XX Homo sapiens.
OS
XX US2002028508-A1.
PN
XX

PD 07-MAR-2002.
XX
XX 21-FEB-2001; 2001US-0790264.
XX
XX 23-APR-1998; 9805-0065363.
PR 23-APR-1998; 9805-0065661.
PR 22-JUN-1998; 9805-0102705.
PR 29-JUL-1998; 9805-0124538.
PR 23-APR-1999; 9905-0298531.
PR 22-JUN-1999; 9905-0337930.
PR 29-JUL-1999; 9905-0363650.
XX
XX (HOLT/) HOLTZMAN D A.
PA (GOOD/) GOODEARL A D J.
XX (MCCA/) MCCARTHY S A.
XX
XX Holtzman DA, Goodearl ADJ, McCarthy SA;
XX WPI; 2002-303420/34.
XX
XX Novel TANGO polypeptides and nucleic acid molecules useful as
PT modulating agents in regulating cellular processes and for diagnosing
PT and treating heart, liver, lung, kidney, inflammatory and cellular
PT proliferative disorders
XX
XX Claim 50; Page 51-52; 138pp; English.

CC The invention relates to nucleic acids encoding a variety of proteins
CC human T139 (TANGO-139), T125 (TANGO-125), T110 (TANGO-110), murine T175
CC (TANGO-175), human T175 or murine WDM-2, having diagnostic, preventive,
CC therapeutic and other uses. Polypeptide of the invention has the ability
CC to inhibit a proteinase activity, to modulate cell-cell interactions, and
CC haematopoiesis and the ability to modulate clotting. Polypeptide and
CC polynucleotide of the invention are useful for diagnosing and treating
CC disorders characterised by their aberrant expression or activity. The
CC antibodies are useful as modulating agents in regulating a variety of
CC cellular processes e.g. cell proliferation and/or cell differentiation.
CC TANGO-139 is useful for treating kidney defects such as kidney failure,
CC TANGO-125 is useful in wound healing and for treating cancer. TANGO-110
CC is useful for treating neoplasia, TANGO-177 or WDM-2 is useful for
CC treating cancer, are useful to treat pancreatic disorders, such as
CC pancreatitis, cerebrovascular disease, and tumours, and injury or trauma
CC to the brain. TANGO-125, 110, 175 molecules treat heart disorders, e.g.,
CC ischaemic heart disease, cardiovascular disorders, such as ischaemic
CC heart disease. TANGO-139, 125, 110 and 175 molecules are useful to treat
CC renal (kidney) disorders, such as glomerular disease (e.g., acute and
CC chronic glomerulonephritis), TANGO-175 is useful to treat uterine
CC disorders, hyperplasia of the endometrium. TANGO-110 is useful to treat
CC spleen, e.g., the fetal spleen, associated diseases and disorder. TANGO-
CC 125 treats prostate disorders, such as inflammatory diseases, Crohn's
CC disease and tumours. TANGO-139, 125, 110, 175 or WDM-2 are useful for
CC treating proliferative disorders, inflammatory disorders. TANGO-175, or
CC WDM-2 activity also include apoptotic disorders, rheumatoid arthritis,
CC systemic lupus erythematosus, insulin-dependent diabetes mellitus,
CC immune-related disorders, e.g., immunodeficiency disorders, viral
CC disorders, cell growth disorders, e.g., cancers and inflammatory
CC disorders and apoptotic disorders. The nucleic acids of the invention
CC are used in gene therapy. The present sequence is human T139 mature
XX protein.
XX

Sequence 420 AA:

Query Match 91.6%; Score 2315.5; DB 23; Length 420;
Best Local Similarity 97.2%; Pred. No. 2.3e-167;
Matches 417; Conservative 1; Mismatches 2; Indels 9; Gaps 1;

27 EWPMPOLQOAPMAGALNKRSEFLLSLHNRLRSWVOPPADMRRLDMSLSLAQAOARA 86
|||||
Db 1 EWPMPOLQOAPMAGALNKRSEFLLSLHNRLRSWVOPPADMRRLDMSLSLAQAOARA 60
|||||
Qy 87 ALGCIPTPSLASGLMRTLOVGMNMQLLPAGLASFEVYSLWFAEGORYSHAAGCARNAT 146
|||||
Db 61 ALGCIPTPSLASGLMRTLOVGMNMQLLPAGLASFEVYSLWFAEGORYSHAAGCARNAT 120
|||||

QY	147	CHHYQLVWATTSQQLGCGRHILCSAGQTALAEFYCAVSPGGMNEVNGKTIIPYKKGAMCSL	206
Db	121	CHHYQLVWATTSQQLGCGRHILCSAGQAALAEFYCAVSPGGMNEVNGKTIIPYKKGAMCSL	180
QY	207	CTASVSGCFKAMDHAGGLCEVPRNPRMCSQNHGRINISTCHCHCPREYTGRCYQVRCSL	266
Db	181	CTASVSGCFKAMDHAGGLCEVPRNPRMCSQNHGRINISTCHCHCPREYTGRCYQVRCSL	240
QY	267	OCVHGFRREBECSCVCDIGGGAQCATKVFPHHTCDLRIDGCGEFMSSADITYRRAMK	326
Db	241	OCVHGFRREBECSCVCDIGGGAQCATKVFPHHTCDLRIDGCGEFMSSADITYRRAMK	300
QY	327	CORRGVLAQIKSQKQVODILAFYLGRLTETNEYTDSDFETRNFWIGLTYRTAKDSFPMAT	386
Db	301	CORRGVLAQIKSQKQVODILAFYLGRLTETNEYTDSDFETRNFWIGLTYRTAKDSFPMAT	360
QY	387	GEHQATTSRAFGQPRDHHGLWLSAAGFGNCVELQASAAFNMDQCKTRNRYICQPAQE	446
Db	361	GEHQATTSRAFGQPRDHHGLWLSAAGFGNCVELQASAAFNMDQCKTRNRYICQPAQE	411
QY	447	HISRMWPGS 455	
Db	412	HISRMWPGS 420	
RESULT	11		
ABG12157			
ID	ABG12157	standard; Protein; 506 AA.	
XX			
AC	ABG12157;		
XX			
DT	18-FEB-2002	(first entry)	
XX			
DE	Novel human diagnostic protein #12148.		
XX			
KM	Human; Chromosome mapping; gene mapping; gene therapy; forensic;		
XX	food supplement; medical imaging; diagnostic; genetic disorder.		
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
XX	23-AUG-2000; 2000US-0649167.		
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmnac RT, Liu C, Tang YF;		
XX			
DR	WPI: 2001-639362/73.		
XX	N-PSDB; AAS76344.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
XX	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
XX	biodiversity		
PS	Claim 20; SEQ ID NO 42516; 103pp; English.		
XX			
XX	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridization probes,		
CC	(polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		

CC a food complement (II) and its binding partners are useful in medical
CC imaging of sites expressing (I), (I') and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pol_sequences.

SQ Sequence 506 AA;

Query Match	77.68; Score 1961.5; DB 22; Length 506;
-------------	---

Matches 369; Conservative 12; Mismatches 23; Indels 95; Gaps 7;

02 6 TSPGRGHLAVLLALGTT-----W--AEVWPQLQEQAPMAGALNRKESFLLSL 54

Db 42 SDPGQPVLPAVTSVVGYSCLGPQEGRLRWSLAGSWHP---FQGPVKVL---ENFISVK- 94

55 HNR LRSWVQPPADMRRLDWSDSLAQLAQARAALCGIPTPSIASGLWRTLQVGWNMQLLP 114

```
Db      95 -----DWSDSLAQLAQRALCGTPTPSLASGLWRTLQVGWNMLLP 136
```

QY 115 AGLASFEVVSIMFAEGQRYSHAAGECARNATCTHYTQLVWATSSQLGCGRHLC SAGQTA 174

Db 137 AGLVSFVEVSLWEAGQRYSHAAGECARNATCTHYTQLVWATSSQLGCGRHLC SAGQAA 196

QY	175	IEAFVCAYS	PGGNWEVNGKTI	IPYKKGAWCSLCTASVSGCFKAWDHAGGLCEVPRNCRM	234
----	-----	-----------	--------------	--	-----

Db 197 IEAFVCAYS PRGNWEVNGKTI VPYKKGAWCSLCTASVSGCFKAMDHAGGLCEVPRNPCRM 256

235 SCQNHGRLNISTCHCHCPPGYTGRYCQVRCSLQCVHGRFREEECSVCVDIGYGGAQCATK 294

Db 257 SCQNHGRINISTCHCHCPPGYTGRYCQVRCSLQCVHGRFREECSCVCDIGYGGAQCATK 316

295 VHFPEHTCDLRIDGDCFMVSSEADTYRARMKCQKKGVLAIKSKVQDILAFYLGRL 354

Db 317 VHEPHTCDLRIDGCEMVSSADTYRKARMKCQKKGVLAIQIKSQKVQDILAFYLGRL 376

355	TINEVTDSDFETRNEWI	371
QY		

Db 377 TINEVIDSDEFETRNFWIGTLGHGCCQOMARGPGRHEHMGASASAGKRGQKGSWQQT PGSEWA 436

QY 372 -----GLTYKTAKDSFRWATGEHOAF TSTFAFGQPDNHGLVWLSAAMGFGNCVELQASA 424

D6 437 NLDYPGPGLTYKTAKDSFRWATGEHQAFTSFAFGQPDNH-----GFGNCVELQASA 487

QY	425	AFNWNDQRC	TRNR	YICQF	443
----	-----	-----------	------	-------	-----

Db 488 AFNWNDQRCCKTRNRYICQF 506

RESULT 12

ID ABG27603 standard: protein: 395 AA

XX
AC
ABG27603:

DT 18-FEB-2002 (first entry)
XX

Novel human diagnostic protein #27594.

Human: chromosome mapping: gene mapping: gene therapy: forensic: KW XX

kw food supplement; medical imaging; diagnostic; genetic disorder.
xx

XX
OS
XX
Homo sapiens.

PN WO2001/506/-A2.
XX

PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB: AAS91/90.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 57962; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 395 AA:

Query Match 68.5%; Score 1732; DB 22; Length 395;
Best Local Similarity 97.8%; Pred. No. 3.4e-123;
Matches 308; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 MLHETSPGRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLRS 60
DB 33 MLHETSPGRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLRS 92
QY 61 WQPPADMRRLDMSLSLAQLAQRALCGIPPTSLASGLWRTLOYGWNQQLPAGLASF 120
DB 93 WQPPADMRRLDMSLSLAQLAQRALCGIPPTSLASGLWRTLOYGWNQQLPAGLASF 152
QY 121 VEVVSLMFAGGORYSHAGCARNAATCTHYQLWATSSOLCGGRLHCSAGOTATAPFC 180
DB 153 VEVVSLMFAGGORYSHAGCARNAATCTHYQLWATSSOLCGGRLHCSAGOTATAPFC 212
QY 181 AVSPGNMVEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNPMSCQNG 240
DB 213 AVSPGNMVEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNPMSCQNG 272
QY 241 RLNIISCTCHCPCPGYGRYCOVRCSLQCVHGRFRREECSCVCDIGYGAQCATNHFPH 300
DB 273 RLNIISCTCHCPCPGYGRYCOVRCSLQCVHGRFRREECSCVCDIGYGAQCATNHFPH 332
QY 301 TCDLRIDGDCFNWSS 315
DB 333 TCDLRIDGDCFNWSS 347

RESULT 13
ABG12156
ID ABG12156 standard; Protein: 276 AA.

XX AC ABG12156;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #12147.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB: AAS76343.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 42515; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 276 AA:

Query Match 45.3%; Score 1145; DB 22; Length 276;
Best Local Similarity 93.9%; Pred. No. 6.8e-79;
Matches 215; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 1 MLHETSPGRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLRS 60
DB 33 MLHETSPGRGHLLAVLLALGTTMAEVPPOLOEQAPMAGALNRKESFLLSLHNRLRS 92
QY 61 WQPPADMRRLDMSLSLAQLAQRALCGIPPTSLASGLWRTLOYGWNQQLPAGLASF 120

Db 93 WVOPPADMRRLDMSDSLQALQARAALCGIPTPSLASGLMRTLOVGMNMOILLPAGLASF 152
QY 121 VEVVSLMFAGQRYSHAGECARNATCTHTYQLVWATSSQLCGRHLCGAGOTAEAVC 180
CC ||||||| ||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Cc 153 VEVVSLMFAGQRYSHAGEXARNATCTHTYQLVWATSSQLCGRHLCGAGALAEAVC 212
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 181 AYSFGMVEVNGKTIIPYKKAGWCSLCTASVSGCFKAMDHAGLCEVPR 229
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 213 AYSFGMVEVNGKTIIPYKKAGWCSLCTASVSSLKAMDHARGLLGPR 261
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
RESULT 14
AA70015
ID AAY70015 standard; Protein: 166 AA.
AC AAY70015;
XX
XX
05-JUN-2000 (first entry)
XX
XX
Human Protease and associated protein-9 (PPRG-9).
XX
XX
Protease and associated protein-9; PPRG-9; anti-PPRG antibody;
KW diagnosis; treatment; cell proliferative disorder; cancer; cirrhosis;
KW arteriosclerosis; atherosclerosis; bursitis; hepatitis; immune disorder;
KW AIDS; Addison's disease; adult respiratory distress syndrome; allergy;
KW ankylosing spondylitis; amyloidosis; cystostatic; antiarteriosclerotic;
KW hepatotrophic; antiinflammatory; virucide; antipsoriatic; anti-HIV;
KW antiallergic; immunosuppressive; antidiabetic; antianaemic;
KW neuroprotective; human.
XX
XX
Homo sapiens.
OS
XX
XX
Key Location/Qualifiers
FT Peptide 1..26
FT /label= signal_peptide
FT Protein 27..166
FT /label= Mature_PPRG-9
FT Modified-site 48
FT /note= "Potential Phosphorylation site"
FT Modified-site 119
FT /note= "Potential Phosphorylation site"
XX
XX
WO200009709-A2.
PN
XX
XX
24-FEB-2000.
PD
XX
XX
06-AUG-1999; 99WO-US17818.
PF
XX
XX
10-AUG-1998; 98US-0096114.
PR
XX
XX
11-FEB-1999; 99US-0119766.
PR
XX
XX
(INCY-) INCYTE PHARM INC.
PA
XX
XX
Bandman O, Hillman JL, Baughn MR, Azimzai Y, Guegler KJ;
PI Corley NC, Yue H, Tang YT, Reddy R, Patterson C, Au-Young J;
PI Shih LL, Lu DAM;
XX
XX
WPI: 2000-224346/19.
DR N-PSDB: AAZ50926.
XX
XX
New human proteases, useful for diagnosis, treatment and prevention of
PT cell proliferative disorders such as atherosclerosis -
XX
XX
Claim 1: Page 84-85; 114pp; English.
PS
XX
XX
The present sequence is that of human protease and associated protein-9
CC (PPRG-9), which is expressed in urologic and musculoskeletal tissues.
CC It is encoded by cDNA identified in Incyte clone 998626 derived from
CC KINNTJ01 cDNA library. Anti-PPRG antibodies can be used
CC as therapeutic antagonists, reagents for diagnosis and monitoring
CC diseases and for isolating PPRG. PPRG nucleotide sequence can be used
CC as probe or primer for diagnosis and monitoring of PPRG-related
CC diseases and gene mapping. PPRG can be used in the treatment of cell

CC proliferative disorders like cancer, arteriosclerosis, atherosclerosis,
CC bursitis, cirrhosis and hepatitis, and immune disorders like AIDS,
Cc Addison's disease, adult respiratory distress syndrome, allergies,
CC ankylosing spondylitis and amyloidosis.
XX
SQ Sequence 166 AA:
QY
Query Match 34.9%; Score 882; DB 21; Length 166;
Best Local Similarity 99.4%; Pred. No. 3.2e-59;
Matches 165; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 1 MLHPETSPGRGHLLAVLALIGTWAEVWPQLOEQAPMAGALNRKESFLLSLHNRRS 60
1 MLHPETSPGRGHLLAVLALIGTWAEVWPQLOEQAPMAGALNRKESFLLSLHNRRS 60
QY 61 WVOPPADMRRLDMSDSLQALQARAALCGIPTPSLASGLMRTLOVGMNMOILLPAGLASF 120
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 WVOPPADMRRLDMSDSLQALQARAALCGIPTPSLASGLMRTLOVGMNMOILLPAGLASF 120
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 121 VEVVSLMFAGQRYSHAGECARNATCTHTYQLVWATSSQLCGRH 166
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 VEVVSLMFAGQRYSHAGECARNATCTHTYQLVWATSSQLCGRH 166
CC ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
RESULT 15
ABG17023
ID ABG17023 standard; Protein: 493 AA.
AC ABG17023;
XX
XX
18-FEB-2002 (first entry)
DT
XX
XX
Novel human diagnostic protein #17014.
DE
XX
XX
Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX
Homo sapiens.
OS
XX
XX
WO200175067-A2.
PN
XX
XX
11-OCT-2001.
PD
XX
XX
30-MAR-2001; 2001WO-US08631.
PF
XX
XX
31-MAR-2000; 2000US-0540217.
PR
XX
XX
23-AUG-2000; 2000US-0649167.
PR
XX
XX
(HYSE-) HYSEQ INC.
PA
XX
XX
Drimnac RT, Liu C, Tang YT;
PI
XX
XX
WPI: 2001-639362/73.
DR N-PSDB: AAS81210.
XX
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
Claim 20: SEQ ID No 47382; 103pp; English.
PS
XX
XX
The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 493 AA:

Query Match 29.1%, Score 735; DB 22; Length 493;
Best Local Similarity 48.6%; Pred. NO. 1.5e-47;
Matches 159; Conservative 3; Mismatches 21; Indels 144; Gaps 6;

186 GMEVNGKTTIPYKGMCSICTASVSGCFKAMDHAGLCIEVPRNFCRMSCQNHRLNIS 245
129 GMEVNGKTTIPYKGMCSICTASVSGCFKAMDHAGLC----- 168
246 TCHCHCPGTYGRCVRCSLQCVHGRFREECSCVCDIGYGGAQCATKYHF----- 297
169 -----VCSLQCVHGRFREECSCVCDIGYGGAQCATWNTFVKQLYFE 212
298 PFHTCDLRIDGCFWVSSAD-----TYRRAR----- 324
213 QIKECADPEVGVKGVVSCIVDHRGNITEYQCHQYITKMTAIIISDYRLICGFMDCKNDI 272
325 --MNC-----ORKGVLAQIKSQKQVODILAFYLGR 352
273 NILKCGSIRLGEKVSQLAEKWSRKETMTLLGMVGLQKRGSVLAQIKSQKQVODILAFYLGR 332
353 LETTNEVTDSDFETRNFWI-----GLTYKT 377
333 LETTNEVIDSDPETRNFWIGREHMAASAGKRGQKGSWQOTPGSEWANIIDYPGPLTYKT 392
378 AKDSFRWATGEHQAFTSFAFGQPDNHG 404
393 AKDSFRWATGEHQAFTSFAFGQPDNHG 419

Search completed: December 29, 2002, 02:23:00
Job time : 82 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 29, 2002, 02:25:44 ; Search time 72 Seconds
(without alignments)
109.416 Million cell updates/sec

Title: US-09-944-896-50

Perfect score: 2529
Sequence: 1 MLHPETSPGRGHLAVLAL.....ENRYICQFAQHISMGPQS 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 109717 seqs, 17314136 residues

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Published_Applications_AA:*
2: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2529	100.0	455	US-09-944-413-50	Sequence 50, Appl
2	2529	100.0	455	US-09-944-403-50	Sequence 50, Appl
3	2529	100.0	455	US-09-944-896-50	Sequence 50, Appl
4	2529	100.0	455	US-09-944-944-50	Sequence 50, Appl
5	2529	100.0	455	US-09-866-028-50	Sequence 50, Appl
6	2529	100.0	455	US-09-944-449-50	Sequence 50, Appl
7	2529	100.0	455	US-09-944-457-50	Sequence 50, Appl
8	2529	100.0	455	US-09-944-862-50	Sequence 50, Appl
9	2529	100.0	455	US-09-945-587-50	Sequence 50, Appl
10	2529	100.0	455	US-09-945-015-50	Sequence 50, Appl
11	2529	100.0	455	US-09-944-396-50	Sequence 50, Appl
12	2529	100.0	455	US-09-944-097-50	Sequence 50, Appl
13	2529	100.0	455	US-09-944-432-50	Sequence 50, Appl
14	2529	100.0	455	US-09-943-762-50	Sequence 50, Appl
15	2529	100.0	455	US-09-944-654-50	Sequence 50, Appl
16	2529	100.0	455	US-09-943-851A-50	Sequence 50, Appl
17	2447.5	96.8	446	US-10-042-141-47	Sequence 47, Appl
18	2447.5	96.8	446	US-09-726-643-47	Sequence 47, Appl
19	2447.5	96.8	446	US-09-790-264-2	Sequence 2, Appl

20	2315.5	91.6	420	10	US-09-790-264-4	Sequence 4, Appl
21	272	10.8	463	9	US-09-905-291A-285	Sequence 285, App
22	272	10.8	463	9	US-09-902-853-285	Sequence 285, App
23	272	10.8	463	10	US-09-909-320-285	Sequence 285, App
24	272	10.8	463	10	US-09-909-088B-285	Sequence 2, Appl
25	254.5	10.1	255	10	US-09-876-225-2	Sequence 110, App
26	254.5	10.1	258	9	US-09-808-602-110	Sequence 27, Appl
27	241	9.5	300	9	US-09-808-602-27	Sequence 109, App
28	240	9.5	188	9	US-09-808-602-109	Sequence 113, App
29	237.5	9.4	431	9	US-09-808-602-113	Sequence 108, App
30	235.5	9.3	500	9	US-09-808-602-108	Sequence 114, App
31	235.5	9.3	500	9	US-09-808-602-114	Sequence 363, App
32	235.5	9.3	500	9	US-09-808-602-114	Sequence 363, App
33	235.5	9.3	500	9	US-09-978-295A-363	Sequence 363, App
34	235.5	9.3	500	9	US-09-978-697-363	Sequence 363, App
35	235.5	9.3	500	9	US-09-978-192A-363	Sequence 363, App
36	235.5	9.3	500	9	US-09-999-832A-363	Sequence 70, Appl
37	235.5	9.3	500	12	US-10-052-586-70	Sequence 83, App
38	229	9.1	266	10	US-09-876-225-3	Sequence 930, App
39	220	8.7	302	10	US-09-925-301-853	Sequence 2, Appl
40	196	7.8	741	10	US-09-925-301-853	Sequence 95, Appl
41	193.5	7.7	203	10	US-09-832-320-2	Sequence 15, Appl
42	178.5	7.1	1456	9	US-09-870-759-95	Sequence 15, Appl
43	178	7.0	737	9	US-09-944-413-15	Sequence 15, Appl
44	178	7.0	737	9	US-09-944-403-15	Sequence 15, Appl
45	178	7.0	737	9	US-09-944-896-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-09-944-413-50
: Sequence 50, Application US/09944413
: Patent No. US20020156004A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Bolstein, David
: APPLICANT: Eaton, Dan
: APPLICANT: Ferrara, Napoleone
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gerritsen, Mary
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul
: APPLICANT: Grimaldi, Christopher
: APPLICANT: Gurney, Austin
: APPLICANT: Hillan, Kenneth
: APPLICANT: Kijavain, Ivar
: APPLICANT: Napier, Mary
: APPLICANT: Roy, Margaret
: APPLICANT: Tumas, Daniel
: APPLICANT: Wood, William
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACTS ENCODING THE SAME
: FILE REFERENCE: P2548P1C1
: CURRENT APPLICATION NUMBER: US/09/944,413
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: 09/866,028
: PRIOR FILING DATE: 2001-05-25
: PRIOR APPLICATION NUMBER: 60/067,411
: PRIOR FILING DATE: December 3, 1997
: PRIOR APPLICATION NUMBER: 60/069,334
: PRIOR FILING DATE: December 11, 1997
: PRIOR APPLICATION NUMBER: 60/069335
: PRIOR FILING DATE: December 11, 1997
: PRIOR APPLICATION NUMBER: 60/069,278
: PRIOR FILING DATE: December 11, 1997
: PRIOR APPLICATION NUMBER: 60/069,425
: PRIOR FILING DATE: December 12, 1997
: PRIOR APPLICATION NUMBER: 60/069,696
: PRIOR FILING DATE: December 16, 1997
: PRIOR APPLICATION NUMBER: 60/069,694
: PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/221090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: NO. US20020156004A1eember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: NO. US20020156004A1eember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-413-50

Query Match 100.0%; Score 2529; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MLHPETSPGRGLLAVLLALGTTAAEVPPOQOEQAPMAGALNRKESFLLSLHNRLS 60
|||||

Db 1 MLHPETSPGRGLLAVLLALGTTAAEVPPOQOEQAPMAGALNRKESFLLSLHNRLS 60
Qy 61 WQPPAADMRLDWSDSLQALQARALCGIPPPSLASGLMRTLQVGMNQLPAGLASF 120
Db 61 WQPPAADMRLDWSDSLQALQARALCGIPPPSLASGLMRTLQVGMNQLPAGLASF 120
Qy 121 VEVLNMFNEGORYSHAAGECARNACTHYTOLVMAITSSQLGGRHICSAQOTAEFVC 180
Db 121 VEVLNMFNEGORYSHAAGECARNACTHYTOLVMAITSSQLGGRHICSAQOTAEFVC 180
Qy 181 AYSPEGNWEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNDCRSCNHG 240
Db 181 AYSPEGNWEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNDCRSCNHG 240
Qy 241 RLNISTCHCHCPGYTGRCQYRCSLQCYHGRREECSCVCDIGGACQATKVFHPPH 300
Db 241 RLNISTCHCHCPGYTGRCQYRCSLQCYHGRREECSCVCDIGGACQATKVFHPPH 300
Qy 301 TCDLRIDGCEFYVSEADPTYYRARKKCKRGGVLAQIKSKYODILAFYLGRIETTEVNT 360
Db 301 TCDLRIDGCEFYVSEADPTYYRARKKCKRGGVLAQIKSKYODILAFYLGRIETTEVNT 360
Qy 361 DSDEFTRNFWIGLTYKTKADSEFRMATGEHQAFTSFAFGQPDNNGIWLSSAMGFGNCVEL 420
Db 361 DSDEFTRNFWIGLTYKTKADSEFRMATGEHQAFTSFAFGQPDNNGIWLSSAMGFGNCVEL 420
Qy 421 QASAAFNNDQCKTRNRYICQFAOEHSRWGPGS 455
Db 421 QASAAFNNDQCKTRNRYICQFAOEHSRWGPGS 455

RESULT 2

US-09-944-403-50

Sequence 50, Application US/09944403

Patent No. US20020165143A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin

APPLICANT: Botstein, David

APPLICANT: Eaton, Dan

APPLICANT: Ferrara, Napoleone

APPLICANT: Flivaioff, Ellen

APPLICANT: Gerritsen, Mary

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul

APPLICANT: Grimaldi, Christopher

APPLICANT: Gurney, Austin

APPLICANT: Hillan, Kenneth

APPLICANT: Kijavlin, Ivar

APPLICANT: Napier, Mary

APPLICANT: Roy, Margaret

APPLICANT: Tumas, Daniel

APPLICANT: Wood, William

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P2548PICI

CURRENT APPLICATION NUMBER: US/09/944,403

CURRENT FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 09/866,028

PRIOR FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: 60/067,411

PRIOR FILING DATE: December 3, 1997

PRIOR APPLICATION NUMBER: 60/069,334

PRIOR FILING DATE: December 11, 1997

PRIOR APPLICATION NUMBER: 60/069,335

PRIOR FILING DATE: December 11, 1997

PRIOR APPLICATION NUMBER: 60/069,278

PRIOR FILING DATE: December 11, 1997

PRIOR APPLICATION NUMBER: 60/069,425

PRIOR FILING DATE: December 12, 1997

PRIOR APPLICATION NUMBER: 60/069,686

PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,694

PRIOR FILING DATE: December 16, 1997

Db	1	MLHPETSPGCGHLAVLLAVLLDGLTTWEVMPPOLEQOAPMAGLANRKESPLLLSLNHRLS	60
QY	61	WVQPPADMRRIDMDSIALQAAPAAALGCIIPNPISASGIWRTLOYGMNOLLPAGLASF	120
Db	61	WVQPPADMRRIDMDSIALQAAPAAALGCIIPNPISASGIWRTLOYGMNOLLPAGLASF	120
QY	121	VEVYSLFAFGQRYSHAGSCARNATCTHTYQLWMAVTSQLGCGRHLCASGQTAIEAFC	180
Db	121	VEVYSLFAFGQRYSHAGSCARNATCTHTYQLWMAVTSQLGCGRHLCASGQTAIEAFC	180
QY	181	AYSFGGWEEVNGKTIIPYKKGAMCSICTASVSCFKAMDHAGGLCEVPRPNCMSQNHG	240
Db	181	AYSFGGWEEVNGKTIIPYKKGAMCSICTASVSCFKAMDHAGGLCEVPRPNCMSQNHG	240
QY	241	RLNSTCHCPCPYTYRCQVNCSLDCVHGRFREECSCVCDIGYGGAQCATKVFPH	300
Db	241	RLNSTCHCPCPYTYRCQVNCSLDCVHGRFREECSCVCDIGYGGAQCATKVFPH	300
QY	301	TCDLRIDGDCFMVSSSEADITYRARMKCORRGVLAQIKSQKVDILAFYLGRIETNEVT	360
Db	301	TCDLRIDGDCFMVSSSEADITYRARMKCORRGVLAQIKSQKVDILAFYLGRIETNEVT	360
QY	361	DSDETRNFWTGLTYKPAKQSFPMATGEHQAFTSFAGQDNGHGLWLSAAMFGNCVEL	420
Db	361	DSDETRNFWTGLTYKPAKQSFPMATGEHQAFTSFAGQDNGHGLWLSAAMFGNCVEL	420
QY	421	QASAAFWNDORCKTRNRYICQFAQEHISRWPGS	455
Db	421	QASAAFWNDORCKTRNRYICQFAQEHISRWPGS	455
RESULT 3			
US-09-944-896-50			
Sequence 50, Application US/09944896			
Patent No. US20020168715A1			
GENERAL INFORMATION:			
APPLICANT: Baker, Kevin			
APPLICANT: Botstein, David			
APPLICANT: Bacon, Dan			
APPLICANT: Ferrara, Napoleone			
APPLICANT: Filvaroff, Ellen			
APPLICANT: Gerritsen, Mary			
APPLICANT: Goddard, Audrey			
APPLICANT: Godowski, Paul			
APPLICANT: Grimaldi, Christopher			
APPLICANT: Gurney, Austin			
APPLICANT: Hillan, Kenneth			
APPLICANT: Kijavlin, Ivar			
APPLICANT: Napier, Mary			
APPLICANT: Roy, Margaret			
APPLICANT: Tumas, Daniel			
APPLICANT: Wood, William			
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEOTIC			
FILE REFERENCE: P2548P1C1			
CURRENT APPLICATION NUMBER: US/09/944, 896			
CURRENT FILING DATE: 2001-08-31			
PRIOR APPLICATION NUMBER: 09/866, 028			
PRIOR FILING DATE: 2001-05-25			
PRIOR APPLICATION NUMBER: 60/069, 334			
PRIOR FILING DATE: December 11, 1997			
PRIOR APPLICATION NUMBER: 60/069, 335			
PRIOR FILING DATE: December 11, 1997			
PRIOR APPLICATION NUMBER: 60/069, 278			
PRIOR FILING DATE: December 11, 1997			
PRIOR APPLICATION NUMBER: 60/069, 425			

RESULT 3
US-09-944-896--50
Sequence 50, Application US/09944396
Patent No. US20020168715A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Bolstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Fliviaroff, Ellen
APPLICANT: Gerriksen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gunney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kljavin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944, 896
CURRENT FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 09/866, 028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/069, 334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069, 696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 702
PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,286
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020168715A member 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020168715A member 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-896-50

Query Match 100.0%; Score 2529; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 1, 2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MHPETSPGRGLHLLVLLLTGTTAAEVPVPOLOEQAPMAGALNRKESFLLSLHRLRS 60
1 MHPETSPGRGLHLLVLLLTGTTAAEVPVPOLOEQAPMAGALNRKESFLLSLHRLRS 60

QY 61 WQPPADMRRLDMSDSLQAQARAALCGTPPSLASGLMRTIQVGNMOLLPAGLASF 120
Db 61 WQPPADMRRLDMSDSLQAQARAALCGTPPSLASGLMRTIQVGNMOLLPAGLASF 120
QY 121 VEVVSLMFAGQRYSHNAGCARNATCTHYTLQVMATSSQLCGRHLCASQOTAEAFVC 180
Db 121 VEVVSLMFAGQRYSHNAGCARNATCTHYTLQVMATSSQLCGRHLCASQOTAEAFVC 180
QY 181 AYSPPGMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNCRMSQNHG 240
Db 181 AYSPPGMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNCRMSQNHG 240
QY 241 RLNISTCHCHCPGYGRYGVCSLQCVHGRFEESCSCVDIGYGACATVHHPFH 300
Db 241 RLNISTCHCHCPGYGRYGVCSLQCVHGRFEESCSCVDIGYGACATVHHPFH 300
QY 301 TCDLRIDGDFVMSSEADTYRRAMKQKRGVLAQIKSQVODILAFYGLRLETTNEVT 360
Db 301 TCDLRIDGDFVMSSEADTYRRAMKQKRGVLAQIKSQVODILAFYGLRLETTNEVT 360
QY 361 DSDFETRNWIGLTYTAKDSFRWATGEHOAFTSFAGQPDNGLVWLSAAMGFGNCVEL 420
Db 361 DSDFETRNWIGLTYTAKDSFRWATGEHOAFTSFAGQPDNGLVWLSAAMGFGNCVEL 420
QY 421 QASAAFNMDQRCRTNRNYICOPAOEHISRWPGS 455
Db 421 QASAAFNMDQRCRTNRNYICOPAOEHISRWPGS 455

RESULT 4
US-09-944-944-50
Sequence 50, Application US/09944944
Patent No. US20020173463A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavini, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944, 944
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866, 028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067, 411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069, 334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069, 696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 702
PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,066
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020173463A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020173463A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-944-50

Query Match 100.0%; Score 2529; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHAPETSPGRGHLLAVLLALGTTTAEVWPPOLOEQAPMAGALNRKESFLLSLHNRRLS 60
DB 1 MHAPETSPGRGHLLAVLLALGTTTAEVWPPOLOEQAPMAGALNRKESFLLSLHNRRLS 60

QY 61 WVPAPADMRLDMSDSLQALQARALCGIPTPSIASGLWTLQVGNMOLLPAGLASF 120
DB 61 WVPAPADMRLDMSDSLQALQARALCGIPTPSIASGLWTLQVGNMOLLPAGLASF 120
QY 121 VEVSLSFAEGORYSHAGECARNACTHTYOLVWATSSQLCGRHLCSAGTAEAFVC 180
DB 121 VEVSLSFAEGORYSHAGECARNACTHTYOLVWATSSQLCGRHLCSAGTAEAFVC 180
QY 181 AYSFGNWEVNGKTIIPKKGAMCSLCTASVSGCFKAMDHAGLCEVPNPMSCQNHG 240
DB 181 AYSFGNWEVNGKTIIPKKGAMCSLCTASVSGCFKAMDHAGLCEVPNPMSCQNHG 240
QY 241 RLNISTCHCHPFGYGRYCYVRCSLQCVHGRFREECSCVCDIGYGAQCATKYHPPH 300
DB 241 RLNISTCHCHPFGYGRYCYVRCSLQCVHGRFREECSCVCDIGYGAQCATKYHPPH 300
QY 301 TCCLRIDGDCFWSSSEADTYTRARKCQKRGVLAQISQKYODILAFYLGRLTTEVT 360
DB 301 TCCLRIDGDCFWSSSEADTYTRARKCQKRGVLAQISQKYODILAFYLGRLTTEVT 360
QY 361 DSDPSTRFWTGLTYKTAKDSFRWATGEHOAFTSFAFGPDNHLGWLMSAAMFGNCVEL 420
DB 361 DSDPSTRFWTGLTYKTAKDSFRWATGEHOAFTSFAFGPDNHLGWLMSAAMFGNCVEL 420
QY 421 QASAAFNWMDQCKTRNRYICQFAQEHISRMPGS 455
DB 421 QASAAFNWMDQCKTRNRYICQFAQEHISRMPGS 455

RESULT 5
US-09-866-028-50
Sequence 50, Application US/09866028
Patent No. US20020058309A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Batson, Dan
APPLICANT: Ferraro, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gottfredsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kljavin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/866,028
CURRENT FILING DATE: 2001-05-25
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-866-028-50

Query Match 100.0%; Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHAPETSPGRGHLLAVLLALGTTTAEVWPPOLOEQAPMAGALNRKESFLLSLHNRRLS 60
DB 1 MHAPETSPGRGHLLAVLLALGTTTAEVWPPOLOEQAPMAGALNRKESFLLSLHNRRLS 60
QY 61 WVPAPADMRLDMSDSLQALQARALCGIPTPSIASGLWTLQVGNMOLLPAGLASF 120
DB 61 WVPAPADMRLDMSDSLQALQARALCGIPTPSIASGLWTLQVGNMOLLPAGLASF 120

Db 61 WOPPADMRRLDMSDSLQALQARALCGIPTPSLASGLMWTLOVGMNOLLPAGLASF 120
QY 121 VEVYSLMFAEGQRYSHAGSCARNATCTHYTLVWATSSQLCGRHLCASGOTAFAPVC 180
Db 121 VEVYSLMFAEGQRYSHAGSCARNATCTHYTLVWATSSQLCGRHLCASGOTAFAPVC 180
QY 181 AVSPGNMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPNPMSCQNHG 240
Db 181 AVSPGNMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPNPMSCQNHG 240
QY 241 RLNISSCHCHCPGGRYGVRCVGRSLCCVHGRFRFEESCCVCDISYGAQCATKXHPFH 300
Db 241 RLNISSCHCHCPGGRYGVRCVGRSLCCVHGRFRFEESCCVCDISYGAQCATKXHPFH 300
QY 301 TCDLRIDGCFMWSSEADTYRRAMKQKRGVLAQIKSQVODILATFLGLETTEVT 360
Db 301 TCDLRIDGCFMWSSEADTYRRAMKQKRGVLAQIKSQVODILATFLGLETTEVT 360
QY 361 DSDFEFRNFWIGLTYTKAKSFRWATGEHQAFSTFAFGQPDNHGLVLSAAMFGNCVEL 420
Db 361 DSDFEFRNFWIGLTYTKAKSFRWATGEHQAFSTFAFGQPDNHGLVLSAAMFGNCVEL 420
QY 421 QASAFNMWDQCKTRNRYICQFAOEHISRMGPGS 455
Db 421 QASAFNMWDQCKTRNRYICQFAOEHISRMGPGS 455

RESULT 6
US-09-944-449-50
Sequence 50, Application US/09944449
Patent No. US20020102647A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gertlisen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548PICI
CURRENT APPLICATION NUMBER: US/09/944,449
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997

QY 121 VEVYSLMFAEGQRYSHAGSCARNATCTHYTLVWATSSQLCGRHLCASGOTAFAPVC 180
Db 121 VEVYSLMFAEGQRYSHAGSCARNATCTHYTLVWATSSQLCGRHLCASGOTAFAPVC 180
QY 181 AVSPGNMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPNPMSCQNHG 240
Db 181 AVSPGNMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPNPMSCQNHG 240
QY 241 RLNISSCHCHCPGGRYGVRCVGRSLCCVHGRFRFEESCCVCDISYGAQCATKXHPFH 300
Db 241 RLNISSCHCHCPGGRYGVRCVGRSLCCVHGRFRFEESCCVCDISYGAQCATKXHPFH 300
QY 301 TCDLRIDGCFMWSSEADTYRRAMKQKRGVLAQIKSQVODILATFLGLETTEVT 360
Db 301 TCDLRIDGCFMWSSEADTYRRAMKQKRGVLAQIKSQVODILATFLGLETTEVT 360
QY 361 DSDFEFRNFWIGLTYTKAKSFRWATGEHQAFSTFAFGQPDNHGLVLSAAMFGNCVEL 420
Db 361 DSDFEFRNFWIGLTYTKAKSFRWATGEHQAFSTFAFGQPDNHGLVLSAAMFGNCVEL 420
QY 421 QASAFNMWDQCKTRNRYICQFAOEHISRMGPGS 455
Db 421 QASAFNMWDQCKTRNRYICQFAOEHISRMGPGS 455

QY 121 VEVYSLMFAEGQRYSHAGSCARNATCTHYTLVWATSSQLCGRHLCASGOTAFAPVC 180
Db 121 VEVYSLMFAEGQRYSHAGSCARNATCTHYTLVWATSSQLCGRHLCASGOTAFAPVC 180
QY 181 AVSPGNMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPNPMSCQNHG 240
Db 181 AVSPGNMEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPNPMSCQNHG 240
QY 241 RLNISSCHCHCPGGRYGVRCVGRSLCCVHGRFRFEESCCVCDISYGAQCATKXHPFH 300
Db 241 RLNISSCHCHCPGGRYGVRCVGRSLCCVHGRFRFEESCCVCDISYGAQCATKXHPFH 300
QY 301 TCDLRIDGCFMWSSEADTYRRAMKQKRGVLAQIKSQVODILATFLGLETTEVT 360
Db 301 TCDLRIDGCFMWSSEADTYRRAMKQKRGVLAQIKSQVODILATFLGLETTEVT 360
QY 361 DSDFEFRNFWIGLTYTKAKSFRWATGEHQAFSTFAFGQPDNHGLVLSAAMFGNCVEL 420
Db 361 DSDFEFRNFWIGLTYTKAKSFRWATGEHQAFSTFAFGQPDNHGLVLSAAMFGNCVEL 420
QY 421 QASAFNMWDQCKTRNRYICQFAOEHISRMGPGS 455
Db 421 QASAFNMWDQCKTRNRYICQFAOEHISRMGPGS 455

Db 61 WVPPAADMRLDMSDSLQAQLAQAARALCGIPTPSLASGLMFTLQVGNMOLLPAGLASF 120
QY 121 VEVVSLMFAEGQSRSHAAEGECARNACTHYTQLVWATSSQLCGRHLCASGQTAIEAFVC 180
Db 121 VEVVSLMFAEGQSRSHAAEGECARNACTHYTQLVWATSSQLCGRHLCASGQTAIEAFVC 180
QY 181 AVSPGNNNEVNGKTIIPKKGAMCSLCTASVSGCFKAMHAGGLCEVPRNPMSCMONG 240
Db 181 AVSPGNNNEVNGKTIIPKKGAMCSLCTASVSGCFKAMHAGGLCEVPRNPMSCMONG 240
QY 241 RLNISTCHCHPFGYTGRCVRCSLQCVHGRPREBECSCVDIGGAGOCATKVPFPH 300
Db 241 RLNISTCHCHPFGYTGRCVRCSLQCVHGRPREBECSCVDIGGAGOCATKVPFPH 300
QY 301 TCDLRIDGDCFMVSSADTYRRARKKCGKGVLAQIKSQKODILAFYLGRLTTEVT 360
Db 301 TCDLRIDGDCFMVSSADTYRRARKKCGKGVLAQIKSQKODILAFYLGRLTTEVT 360
QY 361 DSDETRNFWIGLYTKTKAKDSFRMATGEHOATSFAGOPDNHGLVWLSAAMGFCNCVEL 420
Db 361 DSDETRNFWIGLYTKTKAKDSFRMATGEHOATSFAGOPDNHGLVWLSAAMGFCNCVEL 420
QY 421 QASAFNNNDORCKTRNRYICQFAQEHISRMGPGS 455
Db 421 QASAFNNNDORCKTRNRYICQFAQEHISRMGPGS 455

RESULT 7

US-09-944-457-50
; Sequence 50. Application US/09944457
; Patent No. US20020110859A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gettlesen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavich, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,457
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,694
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997

; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020110859A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020110859A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 50
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-457-50

Query Match 100.0%; Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Fred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHETSPGRGHLAVLALIGTTWAEVPPOLQOPAMAGALNKRSEFLLSLHNRLRS 60
Db 1 MLHETSPGRGHLAVLALIGTTWAEVPPOLQOPAMAGALNKRSEFLLSLHNRLRS 60
QY 61 WVPPAADMRLDMSDSLQAQLAQAARALCGIPTPSLASGLMFTLQVGNMOLLPAGLASF 120

Db 61 WVPADMRRLWSDLSLAQLAQRALCGIPTPSLASGLMFTLOVGMNQLPAGIASF 120
OY 121 VEVVSLMFAGGORYSHAAGBCARNACTHYTLVWATSSOLGCRHLCSAGOTAIEAFVC 180
Db 121 VEVVSLMFAGGORYSHAAGBCARNACTHYTLVWATSSOLGCRHLCSAGOTAIEAFVC 180
OY 181 AYSBGMWEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNPMSCQNHG 240
Db 181 AYSBGMWEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPRNPMSCQNHG 240
OY 241 RLNTSTHCHCPRPYTGRYCCVRCSLQCVHGFREBECCVCDIGGAQCAKVPFPH 300
Db 241 RLNTSTHCHCPRPYTGRYCCVRCSLQCVHGFREBECCVCDIGGAQCAKVPFPH 300
OY 301 TCCLRIDGDFMVSSEADTYRARMKCKORSGVLAIQISQKQODILAFYLGRLFTTNEVT 360
Db 301 TCCLRIDGDFMVSSEADTYRARMKCKORSGVLAIQISQKQODILAFYLGRLFTTNEVT 360
OY 361 DSPETRFNFIGLTYKRAKDSFRNATGEOAFTSFAFQPDNHLVWLSAAMGFNCVEL 420
Db 361 DSPETRFNFIGLTYKRAKDSFRNATGEOAFTSFAFQPDNHLVWLSAAMGFNCVEL 420
OY 421 QASAFNMWMDORCKTRNNRYTCQFAOEHSIRMGPGS 455
Db 421 QASAFNMWMDORCKTRNNRYTCQFAOEHSIRMGPGS 455

RESULT 8

US-09-944-862-50
Sequence 50, Application US/09944862
Patent No. US20020115145A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Feltara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Geriltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944, 862
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866, 028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067, 411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069, 334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069, 425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069, 696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069, 870
PRIOR FILING DATE: December 17, 1997

PRIOR APPLICATION NUMBER: 60/069, 873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068, 017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070, 440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074, 086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074, 092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075, 945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112, 850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113, 296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146, 222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216, 021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218, 517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254, 311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020115145A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020115145A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
SEQ ID OF SEQ ID NOS: 120
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-862-50

Query Match 100.0%; Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MHPTSPGRGHLAVLLALGTTAEVWPPOLOEOAPVAGALNRKSFLLSLHNRLRS 60
Db 1 MHPTSPGRGHLAVLLALGTTAEVWPPOLOEOAPVAGALNRKSFLLSLHNRLRS 60
OY 61 WVPADMRRLWSDLSLAQLAQRALCGIPTPSLASGLMFTLOVGMNQLPAGIASF 120

Db 61 WVQPADMRRLDWSLSLAQLAQAARALCGIPTPSLASGLMRTLQVGMNQLLPAGIASF 120
Qy 121 VEVVSLMFAEGGRYSHAGAECAARNATCTHYTLQVWATSSQLCGGRHLCSAGOTAIEAFVC 180
Db 121 VEVVSLMFAEGGRYSHAGAECAARNATCTHYTLQVWATSSQLCGGRHLCSAGOTAIEAFVC 180
Qy 181 AVSPGNEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPRNPRMCSQNHG 240
Db 181 AVSPGNEVNGKTIIPYKKGAMCSLCTASVSGCFKAMDHAGGLCEVPRNPRMCSQNHG 240
Qy 241 RLNITCCHCHPCPTGTRGRCQVRCISQCHGRRREBECVCDIGVAGACAKVHPPFH 300
Db 241 RLNITCCHCHPCPTGTRGRCQVRCISQCHGRRREBECVCDIGVAGACAKVHPPFH 300
Qy 301 TCCLRIDGDCFWVSEADPYRARRMKCQKRGVLAQIKSQVODILAFYLGRLTTEVT 360
Db 301 TCCLRIDGDCFWVSEADPYRARRMKCQKRGVLAQIKSQVODILAFYLGRLTTEVT 360
Qy 361 DSDFTFRNFWIGLTKAKDSFRMATGEHOAFTSFAFGQPDNHLVLSAAMGFGNCVEL 420
Db 361 DSDFTFRNFWIGLTKAKDSFRMATGEHOAFTSFAFGQPDNHLVLSAAMGFGNCVEL 420
Qy 421 QASAFNMNDQCKTRNRITCQFAQEHISRMGPGS 455
Db 421 QASAFNMNDQCKTRNRITCQFAQEHISRMGPGS 455

RESULT 9
US-09-945-587-50
Sequence 50, Application US/09945587
Patent No. US20020127643A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548PICI
CURRENT APPLICATION NUMBER: US/09/945,587
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997

PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: No. US20020127643A1ember 30, 1999
PRIOR FILING DATE: No. US20020127643A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020127643A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien

Query Match 100.0%; Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLHPTSPGRHLLAVLLALGTTWAETWPPOLQEQAPMAGALNRRESFLLSLHNRLS 60
Db 1 MLHPTSPGRHLLAVLLALGTTWAETWPPOLQEQAPMAGALNRRESFLLSLHNRLS 60
Qy 61 WVQPADMRRLDWSLSLAQLAQAARALCGIPTPSLASGLMRTLQVGMNQLLPAGIASF 120

Db 61 WOPPADMRRLDWSLSIAQLAQAARALCGIPTPSLASGLMFTLOYGMNMLPAGIASF 120
QY 121 VEVYSLMFAEGORYSHAAGECARNATCTHYTLVWATSSOLCGRHLCASAGTAFIAFVC 180
Db 121 VEVYSLMFAEGORYSHAAGECARNATCTHYTLVWATSSOLCGRHLCASAGTAFIAFVC 180
QY 181 AYSBGMWEVNGKTIIPYKKGAWCSLCTASVSGCFKAMDHAGGLCEVPRNPMSCQNHG 240
Db 181 AYSBGMWEVNGKTIIPYKKGAWCSLCTASVSGCFKAMDHAGGLCEVPRNPMSCQNHG 240
QY 241 RLNIHSTGHCHCPPTRYTCOVCSLQCVHGFREBECSVCVDIGGAGQCATKVPFPH 300
Db 241 RLNIHSTGHCHCPPTRYTCOVCSLQCVHGFREBECSVCVDIGGAGQCATKVPFPH 300
QY 301 TCCLRIDGDFMVSSEADTYRRARKKQKGGVLAQIKSQKQODIAFYLGRLTTEVNT 360
Db 301 TCCLRIDGDFMVSSEADTYRRARKKQKGGVLAQIKSQKQODIAFYLGRLTTEVNT 360
QY 361 DSDPERFNFVIGLTYKTAQKDSFRNATGEOAFTSFAFGQPDNHGLWLSAAMGFGNCVEL 420
Db 361 DSDPERFNFVIGLTYKTAQKDSFRNATGEOAFTSFAFGQPDNHGLWLSAAMGFGNCVEL 420
QY 421 QASAFNMNDORCKTRNRNYTCQFAOEHSIRMGPGS 455
Db 421 QASAFNMNDORCKTRNRNYTCQFAOEHSIRMGPGS 455

RESULT 10

US-09-945-015-50
Sequence 50, Application US/09945015
Patent No. US20020132768A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME

FILE REFERENCE: P2548P1C1

CURRENT APPLICATION NUMBER: US/09/945,015

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: 09/866,028

PRIOR FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: 60/067,411

PRIOR FILING DATE: December 3, 1997

PRIOR APPLICATION NUMBER: 60/069,334

PRIOR FILING DATE: December 11, 1997

PRIOR APPLICATION NUMBER: 60/069,335

PRIOR FILING DATE: December 11, 1997

PRIOR APPLICATION NUMBER: 60/069,278

PRIOR FILING DATE: December 11, 1997

PRIOR APPLICATION NUMBER: 60/069,425

PRIOR FILING DATE: December 12, 1997

PRIOR APPLICATION NUMBER: 60/069,696

PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,694

PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,702

PRIOR FILING DATE: December 16, 1997

PRIOR APPLICATION NUMBER: 60/069,870

PRIOR FILING DATE: December 17, 1997

PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,1017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020132768A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020132768A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-945-015-50

Query Match 100.0%; Score 2529; DB 10; Length 455;

Best Local Similarity 100.0%; Pred. No. 1.2e-172;

Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHPBETSPGRGHLAVLALGTTAAEWPPQLOEAPWAGALNKRSEFLLSIHRRLRS 60
Db 1 MHPBETSPGRGHLAVLALGTTAAEWPPQLOEAPWAGALNKRSEFLLSIHRRLRS 60
QY 61 WOPPADMRRLDWSLSIAQLAQAARALCGIPTPSLASGLMFTLOYGMNMLPAGIASF 120

Db 61 WOPPADMRRLDWSDSLQALQARALCGIPPTSLASGLMRLTQVGMNQLPAGLASF 120
Qy 121 VEVVSLMFABEQRIASHAGECARNACTHYTQLVATSSQLCCGRLCSAGOTATAPFC 180
Db 121 VEVVSLMFABEQRIASHAGECARNACTHYTQLVATSSQLCCGRLCSAGOTATAPFC 180
Qy 181 AYSFGGMVEVNGKTIIPKKAGMCSLCTASVSGCFKAMDHAGLCEVPNPMSCONNG 240
Db 181 AYSFGGMVEVNGKTIIPKKAGMCSLCTASVSGCFKAMDHAGLCEVPNPMSCONNG 240
Qy 241 RLNIStCHCHCPGYTRYCOVRCSLQCVHGRFREBECSVCVDIGYGAOCATKVFPH 300
Db 241 RLNIStCHCHCPGYTRYCOVRCSLQCVHGRFREBECSVCVDIGYGAOCATKVFPH 300
Qy 301 TCCLRIDGDCFWVSEATYTRARKCQKRGVLAQIKSQKQDILAFYLRLETNETV 360
Db 301 TCCLRIDGDCFWVSEATYTRARKCQKRGVLAQIKSQKQDILAFYLRLETNETV 360
Qy 361 DSDETRFMWIGLTKTKKDSFRMATGEHOAFTSFAGOPDNHGLVLSAAMGFCNVEL 420
Db 361 DSDETRFMWIGLTKTKKDSFRMATGEHOAFTSFAGOPDNHGLVLSAAMGFCNVEL 420
Qy 421 QASAFNMNDQCKTRNRYICQFAQEHISRMGPGS 455
Db 421 QASAFNMNDQCKTRNRYICQFAQEHISRMGPGS 455

RESULT 11
US-09-944-396-50
Sequence 50, Application US/09944396
Patent No. US20020132981A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Bolstein, David
APPLICANT: Baton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hilljan, Kenneth
APPLICANT: Kljavin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1C1
CURRENT APPLICATION NUMBER: US/09/944,396
PRIOR APPLICATION NUMBER: 2001-09-26
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997

PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,286
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/216,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020132981A1, December 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020132981A1, December 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Sapien
US-09-944-396-50

Query Match 100.0% Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHPETSPGRHLAVLALLGTTVAEWPPOLQEQAPNAGALNRKESFLLSLHNRLRS 60
Db 1 MHPETSPGRHLAVLALLGTTVAEWPPOLQEQAPNAGALNRKESFLLSLHNRLRS 60
Qy 61 WOPPADMRRLDWSDSLQALQARALCGIPPTSLASGLMRLTQVGMNQLPAGLASF 120

Db 61 WOPPADMRRLDMSDSLQIAQARAALCGITPPSLASGLMFTLOYGNMOLLPAGLASF 120
QY 121 VEVVSLMFAEGORYSHAAEGECARNATCTHYTOLVWATSSQLCGRHLCGACOTAEAFYC 180
Db 121 VEVVSLMFAEGORYSHAAEGECARNATCTHYTOLVWATSSQLCGRHLCGACOTAEAFYC 180
QY 181 AYSFGGMEVNGKTIIPKKAMCSLCTASVSGCKRADHDAGLCEVPRNPMSCQNHG 240
Db 181 AYSFGGMEVNGKTIIPKKAMCSLCTASVSGCKRADHDAGLCEVPRNPMSCQNHG 240
QY 241 RLNISTCHHCPCGYTGNVCVRSLOCVHGRFREECSCVCDIGYGACATKVHFPFH 300
Db 241 RLNISTCHHCPCGYTGNVCVRSLOCVHGRFREECSCVCDIGYGACATKVHFPFH 300
QY 301 TCDLRIDGDCFMVSSADTYTTRARKKCGKGVLAQIKSQVODILAFYLGRIETTNEVT 360
Db 301 TCDLRIDGDCFMVSSADTYTTRARKKCGKGVLAQIKSQVODILAFYLGRIETTNEVT 360
QY 361 DSDERTNFMVIGLYTKAKDSFRMATGHOAFSTFAFGOPDNHGLVMTSAAMGFGNCVEL 420
Db 361 DSDERTNFMVIGLYTKAKDSFRMATGHOAFSTFAFGOPDNHGLVMTSAAMGFGNCVEL 420
QY 421 QASAAEFNMNDORCKTRNRYICQFAOEHSRNGPGS 455
Db 421 QASAAEFNMNDORCKTRNRYICQFAOEHSRNGPGS 455

RESULT 12
US-09-944-097-50
Sequence 50, Application US/09944097
Patent No US2002013675A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Botstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijewski, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548P1
CURRENT APPLICATION NUMBER: US/09/944,097
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,694
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997

QY 1 MIHPTSGRGHLLAVLLALCTTMAEVPPOLEQAPMAGALNRRSEFLLSLNRRLS 60
Db 1 MIHPTSGRGHLLAVLLALCTTMAEVPPOLEQAPMAGALNRRSEFLLSLNRRLS 60
QY 61 WOPPADMRRLDMSDSLQIAQARAALCGITPPSLASGLMFTLOYGNMOLLPAGLASF 120
Db 61 WOPPADMRRLDMSDSLQIAQARAALCGITPPSLASGLMFTLOYGNMOLLPAGLASF 120

Query Match 100.0%; Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 VEVVSLWFAEQRYSHAAGECARNACTHYTOLWMTSSQLCGRHLCASGQTAIEAPVC 180
Db 121 VEVVSLWFAEQRYSHAAGECARNACTHYTOLWMTSSQLCGRHLCASGQTAIEAPVC 180
QY 181 AYSFGNNEVNGKTIIPKKGAMCSLCTASVSGCFRAMPHAGGLCVPPNPMSCONHG 240
Db 181 AYSFGNNEVNGKTIIPKKGAMCSLCTASVSGCFRAMPHAGGLCVPPNPMSCONHG 240
QY 241 RLNISTCHCHCPGYTGRYCOVRCSLQCVHGRFREBECSVCVDIGYGAQCATKVFPEH 300
Db 241 RLNISTCHCHCPGYTGRYCOVRCSLQCVHGRFREBECSVCVDIGYGAQCATKVFPEH 300
QY 301 TCCLRIDBDCWVSEALTYTRARKCKQKGVLAQIKSQYODILAFYLGLETTNEVT 360
Db 301 TCCLRIDBDCWVSEALTYTRARKCKQKGVLAQIKSQYODILAFYLGLETTNEVT 360
QY 361 DSDETRFMFGLTYFKTKDSFRMATGEHOAFTSFAFGOPDNLGLWLSAAMGFGNCVEL 420
Db 361 DSDETRFMFGLTYFKTKDSFRMATGEHOAFTSFAFGOPDNLGLWLSAAMGFGNCVEL 420
QY 421 QASAAFMNMDQRCRTNRNYICQFAQEHISRMGPGS 455
Db 421 QASAAFMNMDQRCRTNRNYICQFAQEHISRMGPGS 455

RESULT 13
US-09-944-432-50
; Sequence 50, Application US/09944432
; Patent No. US20020142419A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Bolstein, David
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gettisen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Grimaldi, Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Hillan, Kenneth
; APPLICANT: Kijavain, Ivar
; APPLICANT: Napier, Mary
; APPLICANT: Roy, Margaret
; APPLICANT: Tumas, Daniel
; APPLICANT: Wood, William
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P2548P1C1
; CURRENT APPLICATION NUMBER: US/09/944,432
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 09/866,028
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/067,411
; PRIOR FILING DATE: December 3, 1997
; PRIOR APPLICATION NUMBER: 60/069,334
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,335
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,278
; PRIOR FILING DATE: December 11, 1997
; PRIOR APPLICATION NUMBER: 60/069,425
; PRIOR FILING DATE: December 12, 1997
; PRIOR APPLICATION NUMBER: 60/069,696
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,684
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,702
; PRIOR FILING DATE: December 16, 1997
; PRIOR APPLICATION NUMBER: 60/069,870
; PRIOR FILING DATE: December 17, 1997
; PRIOR APPLICATION NUMBER: 60/069,873
; PRIOR FILING DATE: December 17, 1997

; PRIOR APPLICATION NUMBER: 60/068,017
; PRIOR FILING DATE: December 18, 1997
; PRIOR APPLICATION NUMBER: 60/070,440
; PRIOR FILING DATE: January 5, 1998
; PRIOR APPLICATION NUMBER: 60/074,086
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/074,092
; PRIOR FILING DATE: February 9, 1998
; PRIOR APPLICATION NUMBER: 60/075,945
; PRIOR FILING DATE: February 25, 1998
; PRIOR APPLICATION NUMBER: 60/112,850
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 60/113,296
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 60/146,222
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: PCT/US98/19330
; PRIOR FILING DATE: September 16, 1998
; PRIOR APPLICATION NUMBER: PCT/US98/25108
; PRIOR FILING DATE: December 1, 1998
; PRIOR APPLICATION NUMBER: 09/216,021
; PRIOR FILING DATE: December 16, 1998
; PRIOR APPLICATION NUMBER: 09/218,517
; PRIOR FILING DATE: December 22, 1998
; PRIOR APPLICATION NUMBER: 09/254,311
; PRIOR FILING DATE: March 3, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: June 22, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: September 15, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28409
; PRIOR FILING DATE: No. US20020142419A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: No. US20020142419A1ember 30, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/28301
; PRIOR FILING DATE: December 1, 1999
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: December 16, 1999
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: February 11, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: February 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: March 2, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: March 30, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: May 22, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/20710
; PRIOR FILING DATE: July 28, 2000
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: December 1, 2000
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: February 28, 2001
; NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 50
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-944-432-50

Query Match 100.0%; Score 2529; DB 10; Length 455;
Best Local Similarity 100.0%; Pred. No. 1; 2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHPETSPGRGHLLAVLLALGTTWAEVWPQLOEQAPMAGALNKRSEFLLSLHNRLRS 60
Db 1 MLHPETSPGRGHLLAVLLALGTTWAEVWPQLOEQAPMAGALNKRSEFLLSLHNRLRS 60
QY 61 WVOPPADMRRLDMSDSLQALQAKAALCGIPYPSLASGLWRTILOYGNMOLLPGLASF 120
Db 61 WVOPPADMRRLDMSDSLQALQAKAALCGIPYPSLASGLWRTILOYGNMOLLPGLASF 120

QY 121 VEVVSLMFAEQRYSHAAGECARNAATCTHYTOLWMTSSQLCGGRHLSAGGTAIEAFVC 180
Db 121 VEVVSLMFAEQRYSHAAGECARNAATCTHYTOLWMTSSQLCGGRHLSAGGTAIEAFVC 180
QY 181 AYSFGNNEVNGKTIIPKKGAMCSLCTASVSGCFRAMDHAGLCEVPNPPCMSCQNHG 240
Db 181 AYSFGNNEVNGKTIIPKKGAMCSLCTASVSGCFRAMDHAGLCEVPNPPCMSCQNHG 240
QY 241 RLNIHSTCHCHCPGYTGRCYQVCSLQCVHGRFREECSCVCDIGYGAQCATKVFPEH 300
Db 241 RLNIHSTCHCHCPGYTGRCYQVCSLQCVHGRFREECSCVCDIGYGAQCATKVFPEH 300
QY 301 TCDLRIDBDCWVSEADTYTTRARKKQKGVLAQISQKVODILAFYLGRLTTEVNT 360
Db 301 TCDLRIDBDCWVSEADTYTTRARKKQKGVLAQISQKVODILAFYLGRLTTEVNT 360
QY 361 DSDETRFNFVGLTYKTAKDSFRMTGEHOAFTSFAFGOPDNHGLWLSAAMGFCNVEL 420
Db 361 DSDETRFNFVGLTYKTAKDSFRMTGEHOAFTSFAFGOPDNHGLWLSAAMGFCNVEL 420
QY 421 QASAAFNWMDQCKTRNRYICQFAQEHISRMGPGS 455
Db 421 QASAAFNWMDQCKTRNRYICQFAQEHISRMGPGS 455

RESULT 15
US-09-944-654-50
Sequence 50, Application US/09944654
Patent No. US20020142959A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Bolstein, David
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleone
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerltsen, Mary
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul
APPLICANT: Grimaldi, Christopher
APPLICANT: Gurney, Austin
APPLICANT: Hillan, Kenneth
APPLICANT: Kijavlin, Ivar
APPLICANT: Napier, Mary
APPLICANT: Roy, Margaret
APPLICANT: Tumas, Daniel
APPLICANT: Wood, William
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P2548PICI
CURRENT APPLICATION NUMBER: US/09/944,654
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 09/866,028
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: 60/067,411
PRIOR FILING DATE: December 3, 1997
PRIOR APPLICATION NUMBER: 60/069,334
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,335
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,278
PRIOR FILING DATE: December 11, 1997
PRIOR APPLICATION NUMBER: 60/069,425
PRIOR FILING DATE: December 12, 1997
PRIOR APPLICATION NUMBER: 60/069,696
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,684
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,702
PRIOR FILING DATE: December 16, 1997
PRIOR APPLICATION NUMBER: 60/069,870
PRIOR FILING DATE: December 17, 1997
PRIOR APPLICATION NUMBER: 60/069,873
PRIOR FILING DATE: December 17, 1997

PRIOR APPLICATION NUMBER: 60/068,017
PRIOR FILING DATE: December 18, 1997
PRIOR APPLICATION NUMBER: 60/070,440
PRIOR FILING DATE: January 5, 1998
PRIOR APPLICATION NUMBER: 60/074,086
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/074,092
PRIOR FILING DATE: February 9, 1998
PRIOR APPLICATION NUMBER: 60/075,945
PRIOR FILING DATE: February 25, 1998
PRIOR APPLICATION NUMBER: 60/112,850
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 60/113,296
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 60/146,222
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: PCT/US98/19330
PRIOR FILING DATE: September 16, 1998
PRIOR APPLICATION NUMBER: PCT/US98/25108
PRIOR FILING DATE: December 1, 1998
PRIOR APPLICATION NUMBER: 09/215,021
PRIOR FILING DATE: December 16, 1998
PRIOR APPLICATION NUMBER: 09/218,517
PRIOR FILING DATE: December 22, 1998
PRIOR APPLICATION NUMBER: 09/254,311
PRIOR FILING DATE: March 3, 1999
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: June 22, 1999
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: September 15, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28409
PRIOR FILING DATE: No. US20020142959A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: No. US20020142959A1ember 30, 1999
PRIOR APPLICATION NUMBER: PCT/US99/28301
PRIOR FILING DATE: December 1, 1999
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: December 16, 1999
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: February 11, 2000
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: February 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: March 2, 2000
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: March 30, 2000
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: May 22, 2000
PRIOR APPLICATION NUMBER: PCT/US00/20710
PRIOR FILING DATE: July 28, 2000
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: December 1, 2000
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: February 28, 2001
NUMBER OF SEQ ID NOS: 120
SEQ ID NO 50
LENGTH: 455
TYPE: PRT
ORGANISM: Homo Saplen
US-09-944-654-50

Query Match 100.0%: Score 2529; DB 10; Length 455;
Best local similarity 100.0%: Pred. No. 1,2e-172;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLHPETSPGRGHLLAVLLALGTTWAEVWPOLQEOAPMAGALNRKESFLLSLHNRLS 60
Db 1 MLHPETSPGRGHLLAVLLALGTTWAEVWPOLQEOAPMAGALNRKESFLLSLHNRLS 60
QY 61 WNOPPADMRRLDMSDSLQALQARALCGIPRTSLASGLMRTIQVGNWOLLPRGLASF 120
Db 61 WNOPPADMRRLDMSDSLQALQARALCGIPRTSLASGLMRTIQVGNWOLLPRGLASF 120

```

QY 121 VEYVSLMFAEGORSHAGECARNATCTHYTQLVWATSSQLGGRHLCSAGOTATIEAFVC 180
DB 121 VEYVSLMFAEGORSHAGECARNATCTHYTQLVWATSSQLGGRHLCSAGOTATIEAFVC 180
QY 181 AYSFGNMEVNGKTIIPYKKGAWCSLCTASVSGCFKAMDHAGLCEVPRNCRMSCQNHG 240
DB 181 AYSFGNMEVNGKTIIPYKKGAWCSLCTASVSGCFKAMDHAGLCEVPRNCRMSCQNHG 240
QY 241 RLNIISTCHCHCPGTYTGRYCOVRCSLQCVHGRFREBECCSVCDIGYGACATKXVHEPH 300
DB 241 RLNIISTCHCHCPGTYTGRYCOVRCSLQCVHGRFREBECCSVCDIGYGACATKXVHEPH 300
QY 301 TCCLRIDGDCFMVSSSEADTYRRARMKCORRKGVLAKQIKSOKYQDILAFYLRLETTNEVT 360
DB 301 TCCLRIDGDCFMVSSSEADTYRRARMKCORRKGVLAKQIKSOKYQDILAFYLRLETTNEVT 360
QY 361 DSDFETRNFWIGLTYKTAKDSFRWATGEHQAFTSFAGOPDNHGLVWLSAMGFGNCVEL 420
DB 361 DSDFETRNFWIGLTYKTAKDSFRWATGEHQAFTSFAGOPDNHGLVWLSAMGFGNCVEL 420
QY 421 QASAFNMNDORCTRNRYTCQFAQEHISRMGPGS 455
DB 421 QASAFNMNDORCTRNRYTCQFAQEHISRMGPGS 455

```

Search completed: December 29, 2002, 02:36:08
 Job time : 74 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 2002, 01:55:48 ; Search time 37 Seconds

(without alignments)
510.047 Million cell updates/sec

Title: US-09-944-896-50

Perfect score: 2529

Sequence: 1 MLHPETSPGRGHLAVLLAL.....RNRICQFAQHISRWPGS 455

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

112892

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	252	10.0	245	CRS3_HUMAN	P54108 homo sapien
2	240	9.5	244	TPX1_CAVPO	Q04777 cavia porce
3	236	9.3	1268	PGCN_MOUSE	P55066 mus musculu
4	231	9.1	266	GLIP_HUMAN	P48060 homo sapien
5	228	9.0	245	CRS3_HORSE	Q19010 equus cabal
6	226	8.9	242	HELO_HELHO	Q91055 heloderma h
7	224	8.9	312	YR81_CAEEL	Q09566 caenorhadi
8	221	8.7	243	TPX1_HUMAN	P16562 homo sapien
9	213.5	8.4	249	CRS1_HUMAN	P54107 homo sapien
10	212	8.4	1257	PGCN_MOUSE	P55067 rattus norv
11	209	8.3	244	AEGL_MOUSE	Q03401 mus musculu
12	200	7.9	2738	PGCV_RAT	Q62059 mus musculu
13	200	7.9	3358	PGCV_MOUSE	P13611 homo sapien
14	196	7.8	3358	PGCV_HUMAN	P16563 mus musculu
15	193.5	7.7	243	TPX1_MOUSE	Q09093 gallus gall
16	191	7.6	3562	PGCV_CHICK	P81282 bos taurus
17	190	7.5	3381	PGCV_BOVIN	P12020 rattus norv
18	188.5	7.5	246	AEGL_RAT	P35780 polistes fu
19	187.5	7.4	205	VA5_POLFU	P33154 arabisopsi
20	185.5	7.3	161	PRL_ARATH	P79845 tirimerustu
21	184.5	7.3	183	CRVP_TRIMU	P35782 vespa crabr
22	184	7.3	202	VA52_VESCR	P04284 lycopersico
23	182.5	7.2	159	PR06_LYCES	P1657 vespa manda
24	182	7.2	202	VA5_VESMA	Q04108 lycopersico
25	180.5	7.1	159	PR04_LYCES	P35759 polistes ex
26	180.5	7.1	205	VA5_POLEX	Q05109 polistes an
27	180.5	7.1	209	VA5_POLAN	P81656 polistes do
28	179	7.1	206	VA5_POLDO	P22897 homo sapien
29	178.5	7.1	1456	MANR_HUMAN	P35781 vespa crabr
30	178	7.0	202	VA51_VESCR	Q09054 brachydania
31	176.5	7.0	1213	JAG3_BRARE	P09042 nicotiana t
32	174	6.9	168	PRIC_TOBAC	P35783 vesputia fla
33	174	6.9	204	VA5_VESFL	

34	173	6.8	883	1	PCGB_RAT	P55068 rattus norv
35	171	6.8	168	1	PRIA_TOBAC	P08299 nicotiana t
36	170	6.7	175	1	PRIA_LYCES	Q08697 lycopersico
37	170	6.7	2415	1	PCGA_HUMAN	P16112 homo sapien
38	169	6.7	204	1	VA5_VESMC	P35760 vesputia mac
39	168.5	6.7	912	1	PCGB_BOVIN	Q28062 bos taurus
40	168.5	6.7	2321	1	NTR3_HUMAN	Q9um47 homo sapien
41	168	6.6	204	1	VA5_VESGE	P35784 vesputia ger
42	167	6.6	167	1	PRMS_MAIZE	Q00008 zea mays (m
43	166	6.6	227	1	VA5_VESVU	Q05110 vesputia vul
44	165.5	6.5	799	1	ITBN_DROME	Q27591 drosophila
45	165	6.5	168	1	PRIB_TOBAC	P07053 nicotiana t

ALIGNMENTS

RESULT 1

ID	CRS3_HUMAN	STANDARD:	PRT:	245 AA.
AC	P54108; Q15512;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Cysteine-rich secretory protein-3 precursor (CRISP-3) (SGP28 protein).			
GN	CRISP3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96270732; PubMed=8665901;			
RA	Kretzschmar J., Haendler B., Eberspacher U., Roosterman D.,			
RA	Donner P., Schlenning W.-D.;			
RT	"The human cysteine-rich secretory protein (CRISP) family. Primary			
RT	structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3";			
RL	Eur. J. Biochem. 236:827-836(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96186934; PubMed=8601434;			
RA	Kjeldsen L., Cowland J.B., Johnson A.H., Borregaard N.;			
RT	"SGP28, a novel matrix glycoprotein in specific granules of human			
RT	neutrophils with similarity to a human testis-specific gene product			
RT	and a rodent sperm-coating glycoprotein.";			
RL	FEBS Lett. 380:246-250(1996).			
CC	-1- SUBCELLULAR LOCATION: SECRETED; IN NEUTROPHILS, LOCALIZED IN			
CC	SPECIFIC GRANULES.			
CC	-1- TISSUE SPECIFICITY: SALIVARY GLAND, PANCREAS AND PROSTATE >			
CC	EPIDIDYMIS, OVARY, THYMUS AND COLON.			
CC	-1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;			
CC	INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
CC	EMBL: X95240; CAA64527.1; -			
CC	EMBL: X94323; CAA63984.1; -			
CC	HSSP: P04284; ICPE.			
CC	InterPro: IPR001283; Allrgn_V5/TPX1.			
CC	Pfam: PF00188; SCP: 1.			
CC	PRINTS: PR00837; V5TPXLIKE.			
CC	ProDom: PD000542; Allrgn_V5/TPX1: 1.			
CC	SMART: SM00198; SCP: 1.			
CC	PROSITE: PS01009; SCP_AG5_PRI-SC7_1; 1.			
CC	PROSITE: PS01010; SCP_AG5_PRI-SC7_2; 1.			
CC	GlycoProtein: Signal; Multigene family; Polymorphism.			
CC	SIGNAL 1 20 POTENTIAL.			

FT CHAIN 21 245 CYSTEINE-RICH SECRETORY PROTEIN-3.
 FT CARBOHYD 239 239 N-LINKED (GLCNAC.
 FT VARIANT 106 106 S -> P (IN DBSNP:495335).
 FT VARIANT 134 134 /FTID=VAR_011718.
 FT VARIANT 134 134 A -> S (IN DBSNP:1864312).
 FT VARIANT 134 134 /FTID=VAR_011719.
 SQ SEQUENCE 245 AA; 27630 MW; B4DD79CB7AE9E5F9 CRC64;
 Query Match 10.0%; Score 252; DB 1; Length 245;
 Best Local Similarity 27.1%; Pred. No. 2.2e-13;
 Matches 76; Conservative 44; Mismatches 108; Indels 52; Gaps 12;
 QY 13 LLAVALLALGTTMAEVPD-POLQOAPMAGALNKRSEFL--LTLNRLRSWVOPPA 67
 DB 3 LEPVLEFLV---AGLSPFANEDKDPATLTLTQTVQGEIYVKNHLELRASVPPAR 58
 QY 68 DMRLDMSDSLQALQAAALCGI---PTPSLASGLWRTLOVGMNQLPAGLASFEV 123
 DB 59 NMLKEMWKEAANAOKKAWANOCNRRSNPKDRMTS---LKGCEWLYSSAS-SSMSQA 112
 QY 124 VSLFAEGQRYSHAAGECARATCTHYQLVWATSSQIGCGRHLCASGTALEAFVCAVS 183
 DB 113 IQSFDEYNDEDFGVGKPTPAVYGHYTVWYSSYLVCGNAYCPNQVLYKYVQCYC 172
 QY 184 PGGWEEVNGKTIIPYKKGAWCSLCTASVSGCFKAWDHAGLCEVPRNPMSCQNHGRIN 243
 DB 173 PAGMMA--NRLYVEYEGAPCASCPCND-----DGLC--TNGCK----- 208
 QY 244 ISYCHCPCPGYTGRCYQVRCSCLOVHGRFEE-ECSCVC 282
 DB 209 YEDLYSNCK-----SLKLTITCKHQLVDRSCASCNC 240
 RESULT 2
 ID TPX1_CAVPO STANDARD; PRT: 244 AA.
 AC 060477;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Testis-specific protein TPX-1 precursor (Autoantigen 1) (25 kDa
 DE acrosomal autoantigen) (AAL).
 GN TPX1.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
 ON NCBI_TaxID=10141;
 RX STRAIN=Hartley; TISSUE=Testis;
 RP MEDLINE=96354287; PubMed=9115720;
 RA Foster J.A., Gerton G.L.;
 RA "Autoantigen 1 of the guinea pig sperm acrosome is the homologue of
 RA mouse TPX-1 and human TPX1 and is a member of the cysteine-rich
 RA secretory protein (CRISP) family.";
 RL Mol. Reprod. Dev. 44:221-229(1996).
 RN [2]
 RP SEQUENCE OF 22-41.
 RX MEDLINE=88193219; PubMed=3282555;
 RA Hardy D.M., Huang T.T.F. Jr., Driscoll W.J., Tung K.S.K., Wild G.C.;
 RA "Purification and characterization of the primary acrosomal
 RA autoantigen of guinea pig epididymal spermatozoa.";
 RL Biol. Reprod. 38:423-437(1988).
 CC -1- TISSUE SPECIFICITY: TESTIS.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL: U35712; AAC52616.1;
 CC InterPro: IPR001283; Allrgn_V5/TPX1.
 CC Pfam: PF00188; SCP_1.
 CC PRINTS: PR00837; V5TPX1IKE.
 CC ProDom: PD000542; Allrgn_V5/TPX1.1.
 CC SMART: SM00198; SCP_1.
 CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 CC KEGG: Testis; signal.
 FT SIGNAL 1 21
 FT CHAIN 22 244 TESTIS-SPECIFIC PROTEIN TPX-1.
 SQ SEQUENCE 244 AA; 27248 MW; 58DBDE6ECC16A12 CRC64;
 Query Match 9.5%; Score 240; DB 1; Length 244;
 Best Local Similarity 27.6%; Pred. No. 2e-12;
 Matches 68; Conservative 37; Mismatches 99; Indels 42; Gaps 9;
 QY 51 LSLHNLRSWVOPPAADMRRLDMSDSLQALQAAALCGIPTPSLASGLWRTLOVGMN 110
 DB 41 IINKHNLRSKSVTPPASPAMLMKEMWREAAVNAOKWANRCTL-VHSNPDRTKTSKCGENL 99
 QY 111 QLLPAGLASFEVYSLFAEGQRYSHAAGECARATCTHYQLVWATSSQIGCGRHLC 170
 DB 100 -YMSDPSMSDAIQSFDESQDFTFGVGRSHNAVYGHYTVWYSSYLVCGNAYCPN 158
 QY 171 GQTALEAFVCAYSBPGWEEVNGKTIIPYKKGAWCSLCTASVSGCFKAWDHAGLCEVPRN 230
 DB 159 QDSLKYVYVQYCPAGNNVYTKNT--PYKQIGPCASCP-----GICE--NG 200
 QY 231 PCRMSCQNHGRINLISYCHCPCPGYTGRCYQVRCSCLOVHGRFEEECSCVCIDIGYGAQ 290
 DB 201 LCTNCEVEDL--SNCE-----SLKWTACEH-QLIVERCKATC-----R 238
 QY 291 CATKVH 296
 DB 239 CEDKIV 244
 RESULT 3
 ID PGCN_MOUSE STANDARD; PRT: 1268 AA.
 AC P55066;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurocan core protein precursor.
 GN CSPG3 OR NCAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX STRAIN=Balb/c; TISSUE=Brain;
 RP MEDLINE=96039250; PubMed=7490074;
 RA Rauch U., Grimpe B., Kulbe G., Arnold-Ammer I., Beier D.,
 RA Faessler R.;
 RA "Structure and chromosomal localization of the mouse neurocan gene.";
 RL Genomics 28:405-410(1995).
 CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
 CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NC-CAM
 CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
 CC ACID.
 CC -1- TISSUE SPECIFICITY: BRAIN.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.


```

CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Produced by the venomous gland.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U13619; AAC59730.1; -.
CC InterPro: IPR001283; Allrgn_V5/TPx1.
CC Pfam: PF00188; SCP.1.
CC PRINTS: PR00837; V5TPXLKE.
CC PRODOM: PD000542; Allrgn_V5/TPx1; 1.
CC SMART: SM00198; SCP.1.
CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; FALSE_NEG.
CC KEGG: Calcium channel inhibitor; Neurotoxin; Toxin; Signal.
CC SIGNAL: 1 19 POTENTIAL.
CC FT CHAIN: 20 242 HELIOTHERMINE.
CC SEQUENCE: 242 AA; 27493 MW; 0E183FC2F925DF3C CRC64;

Query Match 8.9%; Score 226; DB 1; Length 242;
Best Local Similarity 25.6%; Pred. No. 2.7e-11;
Matches 69; Conservative 38; Mismatches 125; Indels 38; Gaps 7;

QY 13 LLAVLALALGTWAEVWPQLOEQAPMACALNKESEFLSLHNRLRSWVQPPAADMRRL 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 6 LYLCAAMHGHSEGEASP-----KLPGMTSNPDQOTEIDKNNLRIRIYEPASNNLKM 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 DMSDSLALQARALCGIPTPSLASGLMRTLOYGNMQLPGLASFVEVYSLMFEQ 132
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 TWSNITIAQNAQNSANOCSTLEHTSKBERTIDVBCGENTL-FFSSAPYWSYAIONWFERK 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 133 RYSHAAGECARNATCTHYQLVWATSSQLGCGRHLCASAGOTAEAFVCAVSPGNWVNG 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 120 YRFNVGPTAONVMIGHYIYQVWYRSEELGCAIAYCPDPTTYKYQYQYCPGNG--IRS 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 193 KTIIPYKKGAWCSLCTASVSGCFKAMDHAGLCEVPRNRCMQNGLRINISTCHCNP 252
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 RKYTPYSIGPPGDDCDAD-----NGLC--TNPKQ-----NDVYNNCP 215
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 253 PGYTRGYQVRCSLQCVHGRFREECSCVC 282
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 216 -----DLKKVGGCHPIMKDMATCKC 237
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
YR8L_CAEEL STANDARD; PRT; 312 AA.
AC Q09566;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 35.0 kDa protein F48B8.1 in chromosome III.
GN F48B8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kirsten J.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U23514; AAC46538.1; -.
CC WormPep: F48B8.1; CE01953.
CC InterPro: IPR001283; Allrgn_V5/TPx1.
CC Pfam: PF00188; SCP.1.
CC PRINTS: PR00837; V5TPXLKE.
CC PRODOM: PD000542; Allrgn_V5/TPx1; 1.
CC SMART: SM00198; SCP.1.
CC PROSITE: PS01009; SCP_AG5_PRL_SC7_1; FALSE_NEG.
CC PROSITE: PS01010; SCP_AG5_PRL_SC7_2; FALSE_NEG.
CC KEGG: Hypothetical protein.
CC FT DOMAIN: 265 268 POLY-THR.
CC FT DOMAIN: 299 303 POLY-GLU.
CC SEQUENCE: 312 AA; 35054 MW; AEECTBFF25E26288 CRC64;

Query Match 8.9%; Score 224; DB 1; Length 312;
Best Local Similarity 29.7%; Pred. No. 5.1e-11;
Matches 69; Conservative 28; Mismatches 75; Indels 60; Gaps 13;

QY 29 WPPLOEQAPMACALNKE--SFL--LTLHNRLRSWVQPPAADMRRLDMSDLAQLA 82
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 59 FPSHQSD---SGLSRSEHPEYLLKKTTHENRRRV--PASDMNLLYNDELAASA 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 83 QARALCGIPTPSLASGLMRTLOYGNMQLPAGLASFEVYVSLMFEQSR---YSHA 138
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 114 QHADVTCDFRHSR-----GRINVENIWAAP--YSNYSDAISIMENEHNPRCGCNH 165
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 139 GECARNATCTHYQLVWATSSQLGCGRHLCSA-----GOTAEAFVCAVSPGNWVNG 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 166 KHC-----CGHYVQVWAKTNLVGCGFSRCRDYQGVGNGHRAVFCNINPQCN----- 214
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 194 TIIPYKKG-----AW-----CSLCTASVSGCFKAMDHAGLCEVPRN 230
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 TVFVTAARGQLVMPAFWTASGDNKCSNCPANAPACYO-----GLCYMPKN 260
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
TPX1_HUMAN STANDARD; PRT; 243 AA.
AC P16562;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Testis-specific protein TPX-1 precursor (Cysteine-rich secretory
DE protein-2) (Crisp-2).
GN TPX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RX Kaashara M., Gutknecht J., Brew K., Spurr N., Goodfellow P.N.;
RT "Cloning and mapping of a testis-specific gene with sequence
RT similarity to a sperm-coating glycoprotein gene.";
RL Genomics 5:527-534(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RX Kraetzschmar J., Haendler B., Eberspaecher U., Roosterman D.,
RA Donner P., Schleuning W.-D.;
RT "The human cysteine-rich secretory protein (CRISP) family. Primary
RT structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3.";
RL Eur. J. Biochem. 236:827-836(1996).

```

```

CC -1 SUBCELLULAR LOCATION: Secreted (Probable).
CC -1 TISSUE SPECIFICITY: TESTIS AND EPIDIDYMIS.
CC -1 SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FONGI SC7/SC14 AND PLANTS PR-1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M25532; AAA61220.1; -.
CC DR EMBL: X95238; CAA64526.1; -.
CC DR PIR: B33329; B33329.
CC DR Genew: HGNC:12024; TPX1.
CC DR MIM: 187430; -.
CC DR InterPro: IPR001283; Allrgn_V5/TPx1.
CC DR Pfam: PF00188; SCP_1.
CC DR PRINTS: PRO00837; V5TPXLIKE.
CC DR Prodom: PD000542; Allrgn_V5/TPx1; 1.
CC DR SMART: SM00198; SCP_1.
CC DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
CC DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
CC KW Testis; Signal; Multigene family.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 243 TESTIS-SPECIFIC PROTEIN TPX-1.
CC SQ SEQUENCE 243 AA; 27259 MW; C5FE698CA49CFAD9 CRC64;

Query Match 8.7%; Score 221; DB 1; Length 243;
Best local Similarity 28.1%; Pred. No. 6.9e-11;
Matches 72; Conservative 36; Mismatches 88; Indels 60; Gaps 10;

QY 13 LLAVL-----ALLGTWAEVWPOLQEQAPMAGALNKRKESFLLSLHRLR 59
   | : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9 LVTVLLPLPAEKGDPFTALLT-----QLVQRE-----LYNKHHELR 48

QY 60 SWQPRADMRRLDMSLSLAOLAARALCGI---PPPSLASGLWRTQVGNMOLIPA 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 49 KAASPPASNNLKKMSEVYTTNAQRWANKCTLOHSDPEDRKT-----TRCGENTL-YMSS 102

QY 116 GLASFEVVALMPAEGORYSHAGECARNATCTHYTOLVWATNSOLGCGRHLCASQTAI 175
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 103 DPTSSSAIDSWTDELIDFYVGVCPKSNAYVGHITOLVWSTYOVGCGIAYCPNDSLK 162

QY 176 EAFVCAYSPEGNEVNGKTIIPYKGAWS-----LCTAS-----VSGCFKAMID 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 YYYVCGYCPAGN-NMNRKN-TPYQGQGPACCPDCCDKGLCTNSQYQDULMSDCSLKNT 220

QY 221 AGGLCEVPNRNPMSC 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 AGCEHELLEKCKATC 236

RESULT 9
ID CRSL_HUMAN STANDARD; PRT; 249 AA.
AC P54107; Q13248; O00698; Q14082;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cysteine-rich secretory protein-1 precursor (acidic epididymal
   glycoprotein homolog) (ABG-like protein) (ARP).
OS AEG1 OR CRISP-1.
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND SEQUENCE OF 22-41.
RX MEDLINE=96270732; PubMed=8665901;
RA Kretschmar J., Haendler B., Eberspaecher U., Roostermann D.,

```

```

RA Donner P., Schlenker W.-D.: "The human cysteine-rich secretory protein (CRISP) family. Primary
RT structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3.",
RL Eur. J. Biochem. 236:827-836(1996).
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX MEDLINE=96103955; PubMed=8543280.
RY Hayashi M.:
RZ "Analysis of the human acidic epididymal glycoprotein-like molecule:
RT isolation of cDNA and tissue localization."
RL Hokkaido Igaku Zasshi 70:743-753(1995).
RN [3]
RP SEQUENCE FROM N.A. (LONG ISOFORM), AND CHARACTERIZATION.
RX MEDLINE=96435914; PubMed=8838800.
RY Hayashi M., Fujimoto S., Takano H., Ushiki T., Abe K., Ishikura H.,
RZ Yoshida M.C., Kirchhoff C., Ishibashi T., Kasahara M.:
RT "Characterization of a human glycoprotein with a potential role in
RT sperm-egg fusion: cDNA cloning, immunohistochemical localization,
RL and chromosomal assignment of the gene (AEG1L).".
RN Genomics 32:367-374(1996).
CC -I- FUNCTION: MAY HAVE A ROLE IN SPERM-EGG FUSION AND MATURATION.
CC -I- SUBCELLULAR LOCATION: LOCATED IN THE LUMEN AND EPITHELIUM OF
CC DISTAL DUCTUS EFFERENTES AND EPIDIDYMAL DUCTS, AND ON THE
CC POSTACROSOMAL REGION OF THE SPERM HEAD.
CC -I- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC short form/CRISP-1 delta; may be produced by alternative splicing.
CC -I- TISSUE SPECIFICITY: CAPUT, CORPUS, AND CAUDA REGIONS OF THE
CC EPIDIDYMS, THE DUCTUS DEFERENS, SPERM, AND SEMINAL PLASMA.
CC -I- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/Tpx1,
CC INSECTS AG3/AG5, FUNGI SCT/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X95237; CAA64524.1; -
DR EMBL; X95238; CAA64525.1; -
DR EMBL; D380310; AAB35899.1; -
DR EMBL; D38451; BAA07483.1; -
DR GeneW; HGNC:304; AEBGL1.
DR MIT; 601193; -
DR InterPro; IPR001283; Allrgn_V5/Tpx1.
DR Pfam; PF00188; SCP_1.
DR PRINTS; PRO0837; V5TPXLIKE.
DR ProDom; PD000542; Allrgn_V5/Tpx1_1.
DR SMART; SM00198; SCP_1.
DR PROSITE; PS01009; SCP_AGS_PRI_SCT_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRI_SCT_2; 1.
KW Glycoprotein; Signal; Multigene family; Alternative splicing.
FT SIGNAL 1 21
FT CHAIN 22 249 CYSTEINE-RICH SECRETORY PROTEIN-1.
FT CARBOHYD 230 230 N-LINKED (GLUCNA...).
FT VARSPIC 178 178 E-> D (IN SHORT ISOFORM).
FT VARSPIC 179 249 MISSING (IN SHORT ISOFORM).
FT CONFLICT 22 22 K-> E (IN REF. 2 AND 3).
FT CONFLICT 97 97 T-> A (IN REF. 2 AND 3).
FT CONFLICT 97 97 T-> A (IN REF. 2 AND 3).
SQ SEQUENCE 249 AA; 28481 MW; 85BED3DEAD62A9C CRC64;

Query Match 8.4%; Score 213.5; DB 1; Length 249;
Best Local Similarity 27.2%; Pred. No. 2,8e-10;
Matches 66; Conservative 40; Mismatches 90; Indels 47; Gaps 11;

OY 31 POLOEAPPMAGALNRKESFLLSLHNLRLSVYOPPADMRRLMDSLSLAQLAQARAALCG 90
Db 38 PNVOEE-----IVNHNLRRRVYPASNMLKMSSEDAQNAVARFSKYCD 83
OY 91 IPTPLASGLIMRTLQ---VGNMQLLPAGLASFEVEVSLNFAEQARYSHAAGECAR---N 144
Db 84 M---TENPLERLPNPFGEENHMTGYIPV-SWSYSGVMYSSETSKRH--GEWTTTDDDD 137

```

OY 145 ATCTHTYLTWATSSQLGCGRHLCASAGATAIEAFVCAYSFGN-WEVNGKTIIPYKKGAM 203
 Db 138 ITTHTYLTWATSSQLGCGRHLCASAGATAIEAFVCAYSFGN-WEVNGKTIIPYKKGAM 194
 OY 204 CSLCTASVSGCFKAMDHAGLCE--VPRNCPNMSCNHLNSTCHCHPCPGYGRVCO 261
 Db 195 CEACPSN-----CEDKCTNDC-IYDEYFDCDIQVHYLGLGNHSTTILFC 239
 OY 262 VRC 264
 Db 240 ATC 242
 RESULT 10
 PGCN_RAT STANDARD: PRT: 1257 AA.
 AC P55067;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurocan core protein precursor (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult core glycoprotein].
 GN CSFG3 OR NCAN.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 [1]
 [1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=92406907; PubMed=1326557;
 RA Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
 RT "Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain."; J. Biol. Chem. 267:19536-19547(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=94230574; PubMed=7513709;
 RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K., Margolis R.U., Grumet M.;
 RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the neural cell adhesion molecules Ng-CAM/CD115 and N-CAM, and inhibits neuronal adhesion and neurite outgrowth."; J. Cell Biol. 125:669-680(1994).
 RL J. Cell Biol. 125:669-680(1994).
 CC -I- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC ACID.
 CC -I- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED IN KIDNEY, LUNG, LIVER AND MUSCLE.
 CC -I- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -I- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
 CC -I- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -I- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -I- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -I- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -I- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC DR EMBL; M97161; AAC37679.1; -
 DR HSP; P00740; IEDM.

DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MMC.
 DR InterPro; IPR001304; Lectin_C.
 DR InterPro; IPR000538; Link.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00047; Ig; 1.
 DR Pfam; PF00059; Lectin_C; 1.
 DR Pfam; PF00084; sushi; 1.
 DR Pfam; PF00193; Xlink; 2.
 DR ProDom; PD000918; Link; 2.
 DR SMART; SM00032; CCP; 1.
 DR SMART; SM00034; CLECT; 1.
 DR SMART; SM00179; EGF_CA; 1.
 DR SMART; SM00001; EGF_like; 1.
 DR SMART; SM00409; IG; 1.
 DR SMART; SM00445; LINK; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR PROSITE; PS01241; LINK; 2.
 DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE; PS00411; C-TYPE_LECTIN_2; 1.
 DR GlycoProtex; Hyalurononic acid; Proteoglycan; Immunoglobulin domain; EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 1257
 FT CHAIN 639 1257
 FT DOMAIN 51 146
 FT DOMAIN 176 253
 FT DOMAIN 274 355
 FT DOMAIN 949 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1154
 FT DOMAIN 1155 1213
 FT DISULFID 58 139
 FT DISULFID 181 252
 FT DISULFID 205 226
 FT DISULFID 279 354
 FT DISULFID 303 324
 FT DISULFID 953 964
 FT DISULFID 958 973
 FT DISULFID 975 984
 FT DISULFID 1029 1040
 FT DISULFID 1057 1149
 FT DISULFID 1125 1141
 FT DISULFID 1156 1199
 FT DISULFID 1185 1212
 FT CARBOHYD 121 121
 FT CARBOHYD 339 339
 FT CARBOHYD 737 737
 FT CARBOHYD 944 944
 FT CARBOHYD 967 967
 FT CARBOHYD 1164 1164
 SO SEQUENCE 1257 AA; 135544 MW; 992B33DCEFA19EE1B CRC64;
 Query Match 8.4%; Score 212; DB 1; Length 1257;
 Best Local Similarity 24.8%; Pred. No. 2e-09;
 Matches 70; Conservative 34; Mismatches 134; Indels 44; Gaps 10;
 OY 177 AFVCAVSPGNGNEVNGKTIIPYKKGAMCS--ICTASVSGCFKAMDHAGLCEVPNPNCRM 234
 Db 897 ASVSSSEPTRLMDI-PTSLIPVSLGLDESLKVAASPSGLEGEVEVASQDEPTDPCN 955
 OY 235 SCGNHC---RLNISTCHCHPCPGYGRVCOY-----RCSLQCVHGRREBES--CVCD 283
 Db 956 NPLHGRTGRTNGMTGSCDQAGENCETIDIDCLCS-PCENGGTCTIDEVNGFTICL 1014

QY 284 IGYGACATVHPPFHCDL---RIDGCEMVSSPADYYTAAKMGCGRGVLAQIKSQ 340
 DB 1015 PSYGNCEKDPDT---ECCDRGMHFGCHGYFFAHRAMEDEBERORRAGHILTSVSH 1070
 QY 341 KVGQDILAFYGLRLETTNVTDSDFETRNFGWIGLYTAKTADSEFAWGAEGHAFISFAGQP 400
 DB 1071 EENHFI-----NSFGHENSMTGIGLNDRTVEEDFQWTDNTGLQYENWREKOP 1115
 QY 401 DNHGLVLMISAAMFGCNVLOLQASAFAFNNDQRCRTNRNYICQ 442
 DB 1116 DNF-----FAGGEDCVVMAHENGWMDVPCNVLPLYVCK 1150
 RESULT 11
 AEG1_MOUSE
 ID AEG1_MOUSE STANDARD; PRT; 244 AA.
 AC 003401;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Sperm-coating glycoprotein 1 precursor (SCP 1) (Acidic epididymal glycoprotein 1) (Cysteine-rich secretory protein-1) (Crisp-1).
 GN AEG1 OR AEG-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=93246016; PubMed=1301383;
 RA Mizuki N., Kasahara M.;
 RT "Mouse submandibular glands express an androgen-regulated transcript encoding an acidic epididymal glycoprotein-like molecule."; Mol. Cell. Endocrinol. 89:25-32(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=93307144; PubMed=8319566;
 RA Haendler B., Kratzschmar J., Theuring F., Schleuning W.D.;
 RT "Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/AEG) and the novel related CRISP-3 are expressed under androgen control in the mouse salivary gland."; Endocrinology 133:192-198(1993).
 RL [3]
 RT "FUNCTION: THIS PROTEIN IS SUPPOSED TO HELP SPERMATOZOA UNDERGO FUNCTIONAL MATURATION WHILE THEY MOVE FROM THE TESTIS TO THE DUCTUS DEFERENS."
 CC -I- SUBCELLULAR LOCATION: STORED IN SECRETORY GRANULES OF GRANULAR CONVOLUTED TUBULES CELLS.
 CC -I- TISSUE SPECIFICITY: MAINLY FOUND IN THE CAUDA EPIDIDYMIIS WHERE IT IS SYNTHESIZED BY THE PRINCIPAL CELLS AND SECRETED INTO THE LUMEN. BINDS TO THE HEADS OF SPERMATOZOA. ALSO EXPRESSED IN THE SUBMANDIBULAR GLAND.
 CC -I- DEVELOPMENTAL STAGE: EXPONENTIAL INCREASE BETWEEN DAYS 25 AND 30 AFTER BIRTH.
 CC -I- INDUCTION: By androgens.
 CC -I- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1; INSTRUCTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M92849; AAA37185.1; -;
 DR EMBL: L05359; AAA37460.1; -;
 DR PIR: A49202; A49202.
 DR MGD: MGI:102553; Aeg1.
 DR InterPro: IPR001283; ALLgpn_V5/TPX1.
 DR Pfam: PF00188; SCP_1.

DR PRINTS: PR00837; V5TPXLIKE.
 DR Prodom: PD000542; ALLgpn_V5/TPX1; 1.
 DR SMART: SM00198; SCP: 1
 DR PROSITE: PS01009; SCP_AGS_PRI-SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRI-SC7_2; 1.
 KW Sperm; Glycoprotein; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 244 SPERM-COATING GLYCOPROTEIN 1.
 FT CAROHPD 145 145 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 244 AA; 27679 MW; D00DD0348F85781F CRC64;
 Query Match 8.3%; Score 209; DB 1; Length 244;
 Best Local Similarity 25.7%; Pred. No. 6,4e-10;
 Matches 76; Conservative 44; Mismatches 110; Indels 66; Gaps 15;
 QY 13 LLAVALLLGTTWAEVWPPOLOEQAPMAGLNKRKESFL-----LLSLHNLRSKVOPPA 66
 DB 3 LMLVLEFL-----AALPSPSLDSSOENRLEKSLTRKMSVQEBIVSKHNLRRVSPSG 57
 QY 67 ADMRLDMSDSLALQARALCGIPTPSLASGLMRT--LQVGMNMDLPPGLASFVEVY 124
 DB 58 SPLLMENWYDQVNAQMAADKC---TFSSHPIELRTTNLRGENTL-FMSYILASWSSAI 113
 QY 125 SLMPAEGQRYSHAGECARNACTHYTQLVWATSSQLGCGRHLCASGOTALEAFVCAVSP 184
 DB 114 QGVNNEYKDLTYDVGPKPQDSVVGHYTQVYVWNSFPVACGAEBPKPIRY-YVCHYCP 172
 QY 185 GGNMEVNGKTIIPYKKGAMCSICTASVSGCFKAMDNA-GGLCEVPNRPNCMSCONHGRIN 243
 DB 173 VGNVQ--GRLYPTPTAGECASCP-----DHCEDELCL--TNSC----- 206
 QY 244 ISTCHCHCPGYTGRY--CG-VKCSLOCVHPRPREECSCVCDIGYGACATKRVH 296
 DB 207 -----GHEDKTYNCKYLKMLSCHE-ELLKKGCKRATC-----LCEGKH 244
 RESULT 12
 PGCV_RAT
 ID PGCV_RAT STANDARD; PRT; 2738 AA.
 AC OGERBA: 008592; O88564; O9RIK4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial hyaluronate-binding protein) (GHA) (Fragments).
 GN CPBG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A. (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
 RC STRAIN=Wistar Kyoto;
 RX MEDLINE=99327053; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M., Wright T.N.;
 RT "Versican/Pg-M isoforms in vascular smooth muscle cells";
 RL Atheroscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [2]
 RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
 RC STRAIN=Wistar Kyoto;
 RX MEDLINE=98308094; PubMed=9642104;
 RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y., Margolis R.K., Margolis R.U.;
 RT "Differential regulation of expression of hyaluronan-binding proteoglycans in developing brain: aggrecan, versican, neurocan, and brevican";
 RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
 RN [3]
 RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
 RC TISSUE=Kidney;

MEDLINE=96094159; PubMed=9434070;
 RA Pye C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
 "Proteoglycan expression in the normal rat kidney.";
 RL Nephron 77:461-470(1997).
 RN [4]
 RP SEQUENCE OF 2535-2738 FROM N.A.
 RC STRAIN-Sprague-Dawley; TISSUE=Lung;
 RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
 "Molecular cloning and characterization of two developmentally
 regulated genes in rat lung.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 connecting cells with the extracellular matrix. May take part in
 the regulation of cell motility, growth and differentiation. Binds
 hyaluronic acid.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and
 Vint; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
 but not in glomeruli.
 CC -1- DEVELOPMENTAL STAGE: Disappears after the catilage development
 (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF062402; AAC40166.1; -
 CC EMBL: U75306; AAB51125.1; -
 CC EMBL: AF084544; AAD48544.1; -
 CC EMBL: AF072892; AAC26116.1; -
 CC EMBL: AY007691; AAG16631.1; -
 CC HSSP: P01132; IEPG.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF_CA.
 CC InterPro: IPR003599; Ig.
 CC InterPro: IPR003600; Ig_1like.
 CC InterPro: IPR001304; Lectin_C.
 CC InterPro: IPR000538; Link.
 CC InterPro: IPR000436; Sushi_SCR_CCP.
 CC PRINTS: PR01265; LINKMODULE.
 CC SMART: SM00032; CCP; 1.
 CC SMART: SM00034; CLECT; 1.
 CC SMART: SM00181; EGF; 2.
 CC SMART: SM00179; EGF_CA; 2.
 CC SMART: SM00409; IG; 1.
 CC SMART: SM00410; IG_1like; 1.
 CC SMART: SM00445; LINK; 2.
 CC PROSITE: PS00010; ASX_HYDROXYL; 1.
 CC PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 CC PROSITE: PS00041; C-TYPE_LECTIN_2; 1.
 CC PROSITE: PS00022; EGF_1; 2.
 CC PROSITE: PS01186; EGF_2; 1.
 CC PROSITE: PS01187; EGF_CA; 1.
 CC PROSITE: PS01241; LINK; 2.
 CC GlycoProfile: Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal: Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KM Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 2738 VERSICAN CORE PROTEIN.
 FT NON_CONS 348 349

FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 167 244 LINK 1.
 FT DOMAIN 265 346 LINK 2.
 FT DOMAIN <349 695 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
 DOMAIN).
 FT DOMAIN 696 2431 GAG-BETA.
 FT DOMAIN 2431 2467 EGF-LIKE 1.
 FT DOMAIN 2469 2505 EGF-LIKE 2, CALCIUM-BINDING.
 FT DOMAIN 2518 2632 C-TYPE LECTIN.
 FT DOMAIN 2637 2695 SUSHI.
 FT DISULFID 44 130 BY SIMILARITY.
 FT DISULFID 172 243 BY SIMILARITY.
 FT DISULFID 196 217 BY SIMILARITY.
 FT DISULFID 270 345 BY SIMILARITY.
 FT DISULFID 294 315 BY SIMILARITY.
 FT DISULFID 2435 2446 BY SIMILARITY.
 FT DISULFID 2440 2455 BY SIMILARITY.
 FT DISULFID 2457 2466 BY SIMILARITY.
 FT DISULFID 2473 2484 BY SIMILARITY.
 FT DISULFID 2478 2493 BY SIMILARITY.
 FT DISULFID 2495 2504 BY SIMILARITY.
 FT DISULFID 2511 2522 BY SIMILARITY.
 FT DISULFID 2539 2631 BY SIMILARITY.
 FT DISULFID 2607 2623 BY SIMILARITY.
 FT DISULFID 2638 2681 BY SIMILARITY.
 FT DISULFID 2667 2694 BY SIMILARITY.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC...)
 FT CARBOHYD 330 330 N-LINKED (GLCNAC...)
 FT CARBOHYD 692 692 N-LINKED (GLCNAC...)
 FT CARBOHYD 758 758 N-LINKED (GLCNAC...)
 FT CARBOHYD 802 802 N-LINKED (GLCNAC...)
 FT CARBOHYD 805 805 N-LINKED (GLCNAC...)
 FT CARBOHYD 1257 1257 N-LINKED (GLCNAC...)
 FT CARBOHYD 1435 1435 N-LINKED (GLCNAC...)
 FT CARBOHYD 1633 1633 N-LINKED (GLCNAC...)
 FT CARBOHYD 1660 1660 N-LINKED (GLCNAC...)
 FT CARBOHYD 1684 1684 N-LINKED (GLCNAC...)
 FT CARBOHYD 1738 1738 N-LINKED (GLCNAC...)
 FT CARBOHYD 1848 1848 N-LINKED (GLCNAC...)
 FT CARBOHYD 2004 2004 N-LINKED (GLCNAC...)
 FT CARBOHYD 2409 2409 N-LINKED (GLCNAC...)
 FT CARBOHYD 2711 2711 N-LINKED (GLCNAC...)
 FT CARBOHYD 2721 2721 N-LINKED (GLCNAC...)
 FT CARBOHYD 349 2431 MISSING (IN ISOFORM V3).
 FT VARSPLIC 2697 2738 PSAYORTYSKRYLKNSSVYDINSINTSKHRRMSRMOETR
 FT CONFLICT 2535 2539 R -> RKMSPRKNGOPEFNKY (IN ISOFORM VINT).
 FT SEQUENCE 2738 AA; 300004 MW; 12CA626D58BD8C6A CRC64;
 AEREC -> NSARG (IN REF. 4).
 Query Match 7.9%; Score 200; DB 1; Length 2738;
 Best Local Similarity 27.0%; Pred. No. 4.1e-08;
 Matches 65; Conservative 28; Mismatches 84; Indels 64; Gaps 13;
 QY 224 LCEVPRNPRCMSCONRGLNISTCH-----CHCPGYGRGYOY-----RCSL 266
 DB 2434 LKCT--NPC-----LNGCTCYPTETSYVTCACGYSDGQCELDDEBDCBSNPCRGA 2482
 QY 267 QCVHG--RPREECSCYCDIDGYGAOCATKVNPFHTCDL--RIDGDFVWSEADTY 321
 DB 2483 TCVDGLNTRF-----CLCLSYVALCEODT-----ETCDYGMNHFQGGCYKFFAHRRTWD 2533
 QY 322 RARMKCORRGVLAQIKSRVODILAFYLGRLFTTNEVYDSDEFTRNWGLYTKTAKDS 381
 DB 2534 ADRRECLRGAAHLTSLISHEBQ-----MEVNRV-----GHDYO-----WISLNRKMFPHD 2578
 QY 382 FRWATGHOAFSTAFQOPDNHGLVWLSAAMGFGNCVYELQASAFNMWDQCKTNRRTIC 441
 DB 2579 FRWTDGSALOYENRPNQDPS---FFSAG---EDCVITIHENGQMDVPCNTHLITTC 2631
 QY 442 Q 442
 DB 2632 K 2632

FT DISULFID 3288 3315 BY SIMILARITY.
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 348 348 P -> R (IN ISOFORM V1 AND ISOFORM V3).
 FT VARSPLIC 349 348 MISSING (IN ISOFORM V1).
 FT VARSPLIC 1309 3052 MISSING (IN ISOFORM V2).
 FT VARSPLIC 349 3052 MISSING (IN ISOFORM V3).
 FT CONFLICT 126 126 A -> G (IN REF. 3).
 FT CONFLICT 348 348 MISSING (IN REF. 3).
 FT CONFLICT 1658 1658 I -> T (IN REF. 3).
 FT CONFLICT 1674 1680 TWMSNS -> OPGIOTA (IN REF. 3).
 SO SEQUENCE 3358 AA: 366938 MW: 071B80026BC0762D CRC64:

Query Match 7.9%; Score 200; DB 1; Length 3358;
 Best Local Similarity 27.0%; Pred. No. 5.1e-08;
 Matches 65; Conservative 28; Mismatches 84; Indels 64; Gaps 13;

QY 224 LCEVPRNPGRMSCQNGRLNISTCH-----CHCPRGYGRYQV-----KCSL 266
 DB 3055 LCKT--NPC-----LNGGTCYPTETSYCTCARGYSGDCELPDCHSNPCRGKA 3103
 QY 267 QCVHG--RRREECSCVDIGYGACATKVPFHTCDL--RIDDCFWVSEADTY 321
 DB 3104 TCVDSGNTR-----GLCLPSYVGALCEODT---ELCDGVGMHFGQCYVFYFRRTRWD 3154
 QY 322 RARMKCRKGVLAIQKSKQKVDILAFYLRLETTNEVTSDEFTFRNFWIGLTKTKAKDS 381
 DB 3155 AAREECRLGCAHLTSLISHEQ-----MFINRV-----GHDXO-----WGLINKMPEPHD 3199
 QY 382 FRRATEHCAFTSEAFGQDPNNHGLVWLSAMRGNCVELQASAFNNMDQCKTRNRITC 441
 DB 3200 FRMTDSALQYEMWRPNQDS-----FFSAG---EDCVYIITHENGQNDVPCNTHLYTTC 3252
 QY 442 Q 442
 DB 3253 K 3253

RESULT 14
 PCGV_HUMAN STANDARD; PRT; 3396 AA.
 AC P13611: P20754; Q9UNM5; Q13010; Q13189; Q15123;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein precursor (Large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glia)
 DE hyaluronate-binding protein) (GHAP).
 GN CSPG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE=95105188; PubMed=7528742;
 RX Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human

RT versican gene and functional analysis of its promoter.";
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RN TISSUE=Placenta;
 RC TISSUE=90059882; PubMed=2583089;
 RX Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RL EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RN TISSUE=Glial tumor;
 RC MEDLINE=95105187; PubMed=7806529;
 RX Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 alternative splice variants of human versican.";
 RL J. Biol. Chem. 269:32992-32996(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RN TISSUE=Lung fibroblast;
 RC MEDLINE=88007514; PubMed=2820964;
 RX Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 lectin-like and growth factor-like sequences.";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McPherson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 human chromosome 5 (9q12-9q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RN TISSUE=Brain;
 RC MEDLINE=95181355; PubMed=7876137;
 RX Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 without a chondroitin sulfate attachment in region in mouse and human
 tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RN TISSUE=Aortic smooth muscle;
 RC MEDLINE=99327053; PubMed=10397680;
 RX Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wright T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Biol. 19:1630-1639(1999).
 RN [8]
 RP PARTIAL SEQUENCE.
 RN TISSUE=Brain;
 RC MEDLINE=89174663; PubMed=2466833;
 RX Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial
 hyaluronate-binding protein.";
 RL J. Biol. Chem. 264:5981-5987(1989).
 RN [9]
 RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE=96213482; PubMed=8627343;
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -i- FUNCTION: May play a role in intercellular signaling and in
 connecting cells with the extracellular matrix. May take part in
 the regulation of cell motility, growth and differentiation. Binds
 hyaluronic acid.
 CC -i- SUBCELLULAR LOCATION: secreted; extracellular matrix.
 CC -i- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
 CC V2, V3 and Vint; are produced by alternative splicing.
 CC -i- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 in normal brain, gliomas, medulloblastomas, schwannomas,

neurofibromas, and meningiomas; v2 is restricted to normal brain and gliomas; v3 is found in all these tissues except medulloblastomas.

CC - DEVELOPMENTAL STAGE: Disappears after the cartilage development.

CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC - SIMILARITY: CONTAINS 2 LINK DOMAINS.

CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.

CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC - SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

CC - SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

CC -----

CC EMBL: U16306; AAA65018.1; -

CC EMBL: X15998; CAA34128.1; -

CC EMBL: S52488; AAB24878.1; -

CC EMBL: U26555; AAA67565.1; -

CC EMBL: D32039; BAA06801.1; -

CC EMBL: J02814; AAA36437.1; -

CC EMBL: AF084545; AAD48545.1; -

CC PIR: S06014; S06014.

CC PIR: A29348; A29348.

CC PIR: A30358; A30358.

CC HSSP: P01132; 1EGF.

CC GENE: HGNC:2464; CSFG2.

CC MIM: 118661; -

CC InterPro: IPR000152; Asx_hydroxyl.

CC InterPro: IPR000561; EGF-like.

CC InterPro: IPR000742; EGF 2.

CC InterPro: IPR001881; EGF_Ca.

CC InterPro: IPR001438; EGF_II.

CC InterPro: IPR003599; Ig.

CC InterPro: IPR003006; Ig_MHC.

CC InterPro: IPR001304; Lectin_C.

CC InterPro: IPR000538; Link.

CC InterPro: IPR000436; Sush1_SCR_CCP.

CC Pfam: PF00008; EGF; 2.

CC Pfam: PF00047; Ig; 1.

CC Pfam: PF00059; Lectin_C; 1.

CC Pfam: PF00084; sush1; 1.

CC Pfam: PF00193; Xlink; 2.

CC PRINTS: PR00010; EGFBLDOP.

CC ProDom: PD000918; Link; 2.

CC SMART: SM00032; CCP; 1.

CC SMART: SM00034; CLECT; 1.

CC SMART: SM00179; EGF_CA; 1.

CC SMART: SM00001; EGF-like; 1.

CC SMART: SM00409; Ig; 1.

CC SMART: SM00445; Link; 2.

CC PROSITE: PS00010; ASX_HYDROXYL; 1.

CC PROSITE: PS00022; EGF_1; 2.

CC PROSITE: PS01186; EGF_2; 1.

CC PROSITE: PS01187; EGF_CA; 1.

CC PROSITE: PS01241; Link; 2.

CC PROSITE: PS00615; C-TYPE LECTIN_1; 1.

CC PROSITE: PS00615; C-TYPE LECTIN_2; 1.

CC Glycoprotein, Proteoglycan, Lectin, Extracellular matrix; Sush1; Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain; Hyaluronic acid; Alternative splicing.

CC SIGNAL 1 20

CC CHAIN 21 3396

CC DOMAIN 37 137

CC DOMAIN 167 244

CC DOMAIN 265 346

CC DOMAIN 348 1335

CC DOMAIN 1336 3089

CC GAG-BETA.

FT DOMAIN 3089 3125 EGF-LIKE 1

FT DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 3176 3290 C-TYPE LECTIN.

FT DOMAIN 3295 3353 SUSHI.

FT DISULFID 44 130 BY SIMILARITY.

FT DISULFID 172 243 BY SIMILARITY.

FT DISULFID 196 217 BY SIMILARITY.

FT DISULFID 270 345 BY SIMILARITY.

FT DISULFID 294 315 BY SIMILARITY.

FT DISULFID 3093 3104 BY SIMILARITY.

FT DISULFID 3098 3113 BY SIMILARITY.

FT DISULFID 3115 3124 BY SIMILARITY.

FT DISULFID 3131 3142 BY SIMILARITY.

FT DISULFID 3136 3151 BY SIMILARITY.

FT DISULFID 3153 3162 BY SIMILARITY.

FT DISULFID 3169 3180 BY SIMILARITY.

FT DISULFID 3197 3289 BY SIMILARITY.

FT DISULFID 3265 3281 BY SIMILARITY.

FT DISULFID 3296 3339 BY SIMILARITY.

FT DISULFID 3325 3352 BY SIMILARITY.

FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 809 809 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match Best Local Similarity 7.8%; Score 196; DB 1; Length 3396;

Matches 63; Conservative 27; Mismatches 83; Indels 62; Gaps 12;

QY 230 NCRMSQNHGRINLSTCH-----CHCPGYTRYGQV-----RCSLQCVHG- 271

DB 3096 NPC-----LNGSTCYPTSTSYCTCVPGSGDCELDPECHSNPCRNATCVDFG 3146

QY 272 -RPERECCVCDIGYGACQAKVHPFTCDL---RIDGCPMWSSEADTYRARKC 327

DB 3147 NTFR-----CLCLPSYVVALCEQDT---ETCDYGMHKEFGOCYKFFAHRRTWDAEREC 3197

QY 328 QRRGVLAIQIKSQVODIILAFYLGRLTNEVTDSPETNFMWIGLYTKAKDSFRATG 387

DB 3198 RLQGAHLTSLILSHED---MFVNRV-----GHYQ---WIGLDKRMFHDHFRFTDG 3242

QY 388 EHQAFSTFAFGQPDNHLGLVLSAAMGFNCVELQASAAFMNDQCKTRNRYYCQ 442

DB 3243 SYLYQEMRMNPQDS---FFSAG---EDCVIIMHENGQMDVPCVYHLTYLCK 3290

RESULT 15

ID TPX1_MOUSE STANDARD: PRT; 243 AA.

AC P16563;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, last sequence update)

DT 01-FEB-1996 (Rel. 33, last annotation update)

DE Testis-specific protein TPX-1 precursor.

GN TPX1 OR TPX-1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.


```
RC TISSUE-Testis;
RX MEDLINE-90129048; PubMed-2613236;
RA Kasahara M., Gutknecht J., Brew K., Goodfellow P.N.;
RT "Cloning and mapping of a testis-specific gene with sequence
RL similarity to a sperm-coating glycoprotein gene.";
RL Genomics 5:527-534(1989).
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M25533; AAA0472.1; .
CC PIR; A33329; A33329.
CC HSSP; P04284; ICPE.
CC MGD; MGI:98815; TPx1.
CC InterPro; IPR001283; Allrgn_V5/TPx1.
CC Pfam; PF00188; SCP; 1.
CC PRINTS; PR00837; V5TPXLIKE.
CC PRODOM; PD000542; Allrgn_V5/TPx1; 1.
CC SMART; SM00198; SCP; 1.
CC PROSITE; PS01009; SCP_AG5_PRL-SC7_1; 1.
CC PROSITE; PS01010; SCP_AG5_PRL-SC7_2; 1.
CC KW Testis; Signal.
CC FT SIGNAL 1 22 POTENTIAL.
CC FT CHAIN 23 243 TESTIS-SPECIFIC PROTEIN TPX-1.
CC SEQUENCE 243 AA: 27605 MW: 6E707F569ACAA244 CRC64;

Query Match 7.7%; Score 193.5; DB 1; Length 243;
Best Local Similarity 24.0%; Pred. 1.1e-08;
Matches 63; Conservative 30; Mismatches 93; Indels 77; Gaps 11;

QY 51 LLSLNRLRSWYQPPAADRRIDWSDSLAQLAARALCGIPTPS-----LASG----- 99
Db 41 IVNKHNELERRSYNPFGSDILKMEWSIQATTNQKWNKCTLEHSSKDDRRINIRCGENLY 100
QY 100 -----LWRTLLQVGNMQLLPAGLASFEVEVSLFPAEGQRYSHAAGECARNATCTHYTOL 153
Db 101 MSTDPTLMT-----VIOQWYNENEDFYGVG-AKPNASAVGHTQL 140
QY 154 VWAITSQLGCGRHLCASAGTALFAFYCAVSPGQNWENKFTIIPYKGAWCSTCTASVSG 213
Db 141 VWYSSEFKIGCGIAYCPNQDNLEKYFYVCHYCPMGNNVYMKST--PYQGTGTCASCPNN--- 195
QY 214 CFKAMDHAGGLCEVPRNPRMCSQNHGRNLNISTCHCHCPPGYGRYCQVRCSLQCVHGRF 273
Db 196 -----CE--NGLCNNSCDFEDLL--SNCE-----SLKTSAGCKH-EL 227
QY 274 REECCVCVDIGYGAGCATKYH 296
Db 228 LKTKQATC-----LCBDKIH 243

Search completed: December 29, 2002, 02:23:56
Job time : 44 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 2002, 02:21:18 ; Search time 49 seconds
(without alignments)
892.677 Million cell updates/sec

Title: US-09-944-896-50

Perfect score: 2529
Sequence: 1 MLRPETSPGSHLLAVLLAL.....RNRYICQFAQEHISMWPGS 455

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	250	9.9	245	2	S68691 neutrophil granule
2	236	9.3	1268	2	S52781 neutrocan - mouse
3	232	9.2	219	2	JC4131 glioma pathogenesis
4	231	9.1	266	2	JC5308 testis-specific, v
5	224	8.9	312	2	T16415 hypothetical prote
6	221	8.7	243	2	B33329 cysteine-rich secr
7	213.5	8.4	249	2	S68684 cysteine-rich secr
8	212	8.4	1257	2	S28764 neutrocan precursor
9	209	8.3	244	2	A49202 cysteine-rich secr
10	208	8.2	243	2	JE0204 testicular protein
11	200	7.9	2397	1	A55535 versican precursor
12	196	7.8	2409	1	A60979 versican precursor
13	193.5	7.7	243	2	A33329 testis-specific pr
14	191	7.6	3562	2	A47171 chondroitin sulfat
15	190	7.5	1643	2	T14274 versican precursor
16	190	7.5	3381	2	T42389 mannosyl precursor
17	189.5	7.5	1455	1	A48925 mannosyl precursor
18	189	7.5	207	2	T22436 hypothetical prote
19	188.5	7.5	246	2	A24609 acidic epididymal
20	187.5	7.4	205	2	F44583 venom allergen ant
21	185.5	7.3	161	2	J01693 pathogenesis-relat
22	185.5	7.3	212	2	T22437 pathogenesis-relat
23	184	7.3	162	2	T08154 venom allergen ant
24	184	7.3	202	2	H44583 pathogenesis-relat
25	182.5	7.2	159	1	VC0014 pathogenesis-relat
26	181.5	7.2	177	2	T08126 pathogenesis-relat
27	180.5	7.1	159	2	S26238 pathogenesis-relat
28	180.5	7.1	163	2	T04989 pathogenesis-relat
29	180.5	7.1	205	2	A37329 antigen 5 - paper

30	179	7.1	184	2	S10205 pathogenesis-relat
31	178.5	7.1	1456	1	A36563 mannose receptor p
32	178	7.0	202	2	G44583 venom allergen ant
33	177.5	7.0	2825	2	T14271 Doc4 protein, stre
34	177	7.0	178	2	S68681 cysteine-rich secr
35	176.5	7.0	161	2	S65777 pathogenesis-relat
36	176	7.0	185	2	T10677 pathogenesis-relat
37	175	6.9	168	2	T07146 pathogenesis-relat
38	175	6.9	205	2	T48294 pathogenesis-relat
39	175	6.9	208	2	T19852 hypothetical prote
40	174	6.9	168	2	C24620 pathogenesis-relat
41	174	6.9	204	2	A44583 venom allergen ant
42	173.5	6.9	212	2	T27834 hypothetical prote
43	173.5	6.9	213	2	T22439 hypothetical prote
44	173	6.8	209	2	T19848 hypothetical prote
45	173	6.8	883	2	S49126 brevican precursor

ALIGNMENTS

RESULT 1

S68691
neutrophil granules matrix glycoprotein SGP28 precursor - human

C:Species: Homo sapiens (man)

C:Date: 15-Feb-1997 #sequence revision 13-Mar-1997 #text_change 17-Nov-2000

C:Accession: S68691; S74313; S68683

R:Kjeldsen, L.; Cowland, J.B.; Johnsen, A.H.; Borregaard, N.

FEBS Lett. 380, 246-250, 1996

A:Title: SGP28, a novel matrix glycoprotein in specific granules of human neutrophils

A:Reference number: S68691; MUID:96186934; PMID:8601434

A:Accession: S68691

A:Molecule type: mRNA

A:Residues: 1-245 <KUF>

A:Cross-references: EMBL:X94323; NID:g1213612; PIDN:CAA63984.1; PID:g1213613

A:Accession: S74313

A:Molecule type: protein

A:Residues: 33-83;96-143;165-217;221-226 <KUL>

R:Kraetzschmar, J.; Haendler, B.; Eberspaecher, U.; Roosterman, D.; Donner, P.; Schle Eurr. J. Biochem. 236, 827-836, 1996

A:Title: The human cysteine-rich secretory protein (CRISP) family. Primary structure

A:Reference number: S68681; MUID:96270732; PMID:8665901

A:Accession: S68683

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-105; 'S', 107-245 <KRA>

A:Cross-references: EMBL:X95240; NID:g1262818; PIDN:CAA64527.1; PID:g1262819

C:Genetics:

A:Gene: SGP28

C:Superfamily: cysteine-rich secretory protein 1

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-245/Product: neutrophil granules matrix glycoprotein SGP28 #status predicted <MA

Query Match 9.9%; Score 250; DB 2; Length 245;

Best Local Similarity 27.1%; Pred. No. 9.5e-12;

Matches 76; Conservative 44; Mismatches 108; Indels 52; Gaps 12;

QY	13	LLAVLLALIGTMAEWP--POLQDAPMAGALNKRESL---LLSLNRLKRWQPPPA 67
DB	3	LEPVLLFLV---AGLDSPFANEEDKPAFTALLTQTVQREIVNKHHELRVAVSPPAR 58
QY	68	DMRRLDSDSLAQLAARALGCI---PTPSLASGLMPTLQVGMOLLPAGLAFVEV 123
DB	59	NMLKHEMKKEAANQKMANQCNRYNSNKDKMTS-----LKGENDLYMSSAP-SSMSQA 112
QY	124	VSLWAEQORYSHAAGECARNATCTHYTLQVWATSSQLCCGRRHLSAGQTALEAFVCAVS 183
DB	113	IQSWDEYNDFGVGPTPNVAVGHYTVVWYSYLVCAGNAVCPNOKVLKYYVQCXC 172
QY	184	PGGNVEVNGKTIIPYKKAGWCSLCTASVSGCFKAMDHAGGLCEVPNPPRMSCQNGRLN 243
DB	173	PAGNMA--NRLVPEQAGPCASCPDNCD-----DGLC---TNGCK----- 208


```

A:Accession: A29348
A:Molecule type: mRNA
A:Residues: 1725, 'V', 1727-2409 <KRU>
R:Cross-references: GB:J02814
R:Perides, G.; Rahemulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.
J. Biol. Chem. 267, 23883-23887, 1992
A:title: Isolation of a large aggregating proteoglycan from human brain.
A:Reference number: A45131; MUID:93054750; PMID:1429726
A:Contents: brain
A:Accession: A45131
A:Molecule type: protein
A:Residues: 21-22,'X',24-37 <PE2>
A:Experimental source: brain
A>Note: sequence extracted from NCBI backbone (NCBIP:118884)
R:Iozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasnuth, J.J.; McPherson, J.D.
Genomics 14, 845-851, 1992
A:title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human C
A:Reference number: 154179; MUID:93122792; PMID:1478664
A:Accession: 154179
A>Status: translated from GB/EMBL/DBDJB
A:Molecule type: DNA
A:Residues: 251-947 <RES>
A:Cross-references: GB:S52488; NID:g263313; PIDN:AAB24878.L; PID:g263314
C:Genetics:
A:Gene: GDB:CSPG2
A:Cross-references: GDB:127873; OMIM:118661
A:Map position: 5q12-5q14
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;
F.1-20/Domain: signal sequence #status predicted <SIG>
F.21-2409/Product: proteoglycan 24k core protein #status predicted <MAT>
F.167-244/Domain: link protein repeat homology <LNKT>
F.265-346/Domain: link protein repeat homology <LNK2>
F.559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>
F.2106-2137/Domain: EGF homology <EG1>
F.2144-2175/Domain: EGF homology <EG2>
F.2182-2302/Domain: C-type lectin homology <LCH>
F.2309-2365/Domain: complement factor H repeat homology <FHD>

Query Match          7 8%; Score 196; DB 1; Length 2409;
Best Local Similarity 26.8%; Pred. No. 1,1e-06;
Matches 63; Conservative 27; Mismatches 86; Indels 62; Gaps 12;

QY 230 NPCRMSCONHGRINISTCH-----CHCPRGTYGRYCOV-----RCSLOCVHS- 271
      |||              |||||         | ||| | :|         | |||
Db 2109 NPC-----LNGRCPTETSYCTCVPGSGDCGLDFDECHSNPCRGATCVDGF 2159
      |||              |||||         | ||| | :|         | |||

QY 272 -RFREECSCVCICIGTGGAOCATKVHFPHFTCDL--RIDDCENVSSADITYARRMKC 327
      |||              |||||         | ||| | :|         | |||
Db 2160 NTFR-----CLCRPSYVGALCEDOT---ETCDYGWHNFQCQCYKFYHRRRTMDAAEREC 2210
      |||              |||||         | ||| | :|         | |||

QY 328 ORKGVLAIQIKSQKODIIAFYLGRLETFNEVDDPFTNRFMVIGLTAKTADSRFWAG 387
      |||              |||||         | ||| | :|         | |||
Db 2211 RIQGAHLITLSISHEQ----MFNVRV-----GHDIY----WIGLANDRMFHDRFDWDG 2255
      |||              |||||         | ||| | :|         | |||

QY 388 EHQAFTSFAPGCPDNHGWLWSAAMGFCNVELOASAAFNNDORCKTRNNRYICQ 442
      |||              |||||         | ||| | :|         | |||
Db 2256 STLQYENMRPNOPDS----FFSAG---EDCCVILIHENGQMNVDPCNYHLGYTCR 2303
      |||              |||||         | ||| | :|         | |||

RESULT 13
A33329
tests-specific protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Mar-1990 #sequence_reviseon 09-Mar-1990 #text_change 29-Sep-1999
C:Accession: A33329
R:Kasahara, M.; Gutknecht, J.; Brew, K.; Spurr, N.; Goodfellow, P.N.
Genomics 5, 527-533, 1989
A:title: Cloning and mapping of a tests-specific gene with sequence similarity to a
A:Reference number: A33329; MUID:90129048; PMID:2613236
A:Accession: A33329
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-243 <KAS>
```

A:Cross-references: GB:M25533; NID:g202126; PIDN:AAA40472.1; PID:g202127
C:Superfamily: cysteine-rich secretory protein 1

Query Match 7.7%; Score 193.5; DB 2; Length 243;
Best Local Similarity 24.0%; Pred. No. 1.8e-07;
Matches 63; Conservative 30; Mismatches 93; Indels 77; Gaps 11;

QY 51 LLSLHRLRSWQPPADMRRLDMSDSLALQARALCGIPTPS-----LASG----- 99
Db 41 IYNKHNLRRSVNPTGSDILKMEWSIOATNNAOKWANKCILEHSSKDRKINIRCGENLY 100
QY 100 -----LMTLTQYGMWQQLPAGLASFVEVYSLTFEAGORYSHAGECARATCTHYTL 153
Db 101 MSTDTPLTWTST-----VIQSYNENEDFVYGVG-AKPNASVGHYTL 140
QY 154 WATSSQLCGRHLCASGQTAIEAFVCAVSPGMWVNGKTIIPYKKAWSLCTASVSG 213
Db 141 WMYSSFKIGCGIAYCPNQDLKTYFYCHYCPGMNVMKST--PYQGTGPCASCPNN--- 195
QY 214 CRRAMDAGGLCEVPNPPCMSCONHRLNISTCHCHCPRGVGTGRCVRLQCVHGRF 273
Db 196 -----CE--NGLCTNNSCFEDLL--SNCE-----SLKTSAGCKH-EL 227
QY 274 REEECSVCVDIGYGAQCATKVN 296
Db 228 LKTKQCATC-----LCEDKIH 243

RESULT 14
A47171
chondroitin sulfate proteoglycan PG-M core protein - chicken

C:Species: Gallus gallus (chicken)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A47171
R:Shinomura, T.; Ito, K.; Kimata, K.
J. Biol. Chem. 268, 14461-14469, 1993

A:Title: CDNA cloning of PG-M, a large chondroitin sulfate proteoglycan expressed during

A:Reference number: A47171; MUID:93300846; PMID:8314802

A:Accession: A47171

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-3562 <SH1>

A:Cross-references: GB:D13342; NID:g391643; PIDN:BA02742.1; PID:g391644

A:Experimental source: stage 22-23 developing limb buds

A:Note: sequence extracted from NCBI backbone (NCBI:134456, NCBI:134457)

C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lectin

F:166-243/Domain: link protein repeat homology <LNK1>

F:264-345/Domain: link protein repeat homology <LNK2>

F:3258-3289/Domain: EGF homology <EGF1>

F:3296-3327/Domain: EGF homology <EGF2>

F:3334-3454/Domain: C-type lectin homology <LCH>

F:3461-3517/Domain: complement factor H repeat homology <FHD>

Query Match 7.6%; Score 191; DB 2; Length 3562;

Best Local Similarity 25.1%; Pred. No. 4e-06;

Matches 61; Conservative 35; Mismatches 91; Indels 56; Gaps 13;

QY 222 GGLCEVP-RNPPCMS-CONHRLNISTCH-----CHCPRGVGTGRCV----- 262
Db 3247 GTAVQIIPGODPCCKSNPCLNG-----TCYPRGSFYICTLPGFNQCELDIDECOSNPC 3301
QY 263 RGSLOCYHGRREECSCVDCIGYGAQCATKVNHPHTCDL---RIDGCFMWSSEADT 319
Db 3302 RNCATCTIDGL---NTFTCLCLPSTIGALCEODT-----ETCDYGMHKFOGQCYKFFAHRRT 3354
QY 320 YYRAMKRCQKRGVLAQIKSQKVDILAFYLCRLFTTNEVTDSPETRNFWIGLYKTAK 379
Db 3355 WDTAERECRQAGALHTLSHBEQ---VEFNRI-----GHYQ---WIGLNDKME 3399
QY 380 DSRMATGEHOAFTSFAFGQPDNHLGLVLSAAMGFCNVELQASAAFNWDORCKTRNRY 439
Db 3400 RDFRMTGDSPLQYENWRNPDS---FFSAG---EDCVYIWHENGQWMDVPCNHLTY 3452

QY 440 ICQ 442
Db 3453 TCK 3455

RESULT 15
T14274
versican precursor, splice form V2 - bovine

C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000

C:Accession: T14274

R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine b

A:Reference number: Z17954; MUID:96288320; PMID:9624174

A:Accession: T14274

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1643 <SCB>

A:Cross-references: EMBL:AF060458; NID:g3253303; PID:g3253304; PIDN:ANC24360.1

A:Experimental source: brain

C:Keywords: glycoprotein

F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-1643/Product: versican, splice form V2 #status predicted <MAT>
F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (asn) (covalent)

Query Match 7.5%; Score 190; DB 2; Length 1643;
Best Local Similarity 26.0%; Pred. No. 2.2e-06;

Matches 61; Conservative 28; Mismatches 84; Indels 62; Gaps 12;

QY 230 NPCRMSCONHRLNISTCH-----CHCPRGVGTGRCV-----RCSLQCVHG- 271
Db 1343 NPC-----LNGTCTVPTETSYVCTCVPGYSGDRCELDDECHSNPCRNATCIDGF 1393
QY 272 -RREEECSVCVDIGYGAQCATKVNHPHTCDL---RIDGCFMWSSEADTYRRAMKC 327
Db 1394 NTFR-----CLCLPSYVGLCEODT---ETCDYGMHKFOGQCYKFFAHRRTWDAEREC 1444
QY 328 QRKGVLAQIKSQKQVODILAFYLCRLFTTNEVTDSPETRNFWIGLYKTAKDSFRMATG 387
Db 1445 KLQGAHLTSLSHBEQ---MVENRY-----GHYQ---WIGLNDKMEFHFRTWDG 1489
QY 388 EHOAFTSFAFGQPDNHLGLVLSAAMGFCNVELQASAAFNWDORCKTRNRYICQ 442
Db 1490 STLQYENWRNPDS---FFSNG---EDCVYIWHENGQWMDVPCNHYHLYTCK 1537

Search completed: December 29, 2002, 02:26:41
Job time : 55 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 29, 2002, 02:21:48 ; Search time 36 Seconds
(without alignments)
371.873 Million cell updates/sec

Title: US-09-944-896-50
Perfect score: 2529
Sequence: 1 MLHPETSPGRGHLAVLAL.....RNRICQFQPHSHWGPDS 455

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PC105.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	267	10.6	270	2	US-08-773-368-1
2	267	10.6	270	3	US-09-199-887-1
3	232	9.2	219	2	US-08-773-368-3
4	232	9.2	219	3	US-09-199-887-3
5	231	9.1	266	2	US-08-773-368-4
6	231	9.1	266	3	US-09-199-887-4
7	212	8.4	1257	1	US-08-340-4288-49
8	196	7.8	2409	6	5180808-2
9	189.5	7.5	1455	3	US-08-840-062-5
10	185	7.3	151	1	US-08-614-935-28
11	185	7.3	151	3	US-09-130-287-28
12	182.5	7.2	135	2	US-07-857-224B-97
13	182.5	7.2	135	2	US-07-857-224B-98
14	180.5	7.1	135	2	US-07-857-224B-99
15	180.5	7.1	135	2	US-08-419-414-10
16	180.5	7.1	205	1	US-08-614-935-6
17	180.5	7.1	205	1	US-08-614-935-7
18	180.5	7.1	205	3	US-09-130-287-6
19	180.5	7.1	205	3	US-09-130-287-7
20	174	6.9	111	2	US-07-857-224B-105
21	172	6.8	138	2	US-07-857-224B-100
22	172	6.8	163	4	US-09-257-583-13
23	171	6.8	139	2	US-07-857-224B-101
24	171	6.8	168	1	US-08-181-271A-45
25	171	6.8	168	1	US-08-449-315-45
26	171	6.8	168	1	US-08-444-803-45
27	171	6.8	168	1	US-08-449-043-45

28	171	6.8	168	1	US-08-456-265A-45	Sequence 45, Appl
29	171	6.8	168	1	US-08-455-416-45	Sequence 45, Appl
30	171	6.8	168	1	US-08-455-244-45	Sequence 45, Appl
31	171	6.8	168	1	US-08-454-876-45	Sequence 45, Appl
32	171	6.8	168	2	US-08-457-364-45	Sequence 45, Appl
33	171	6.8	168	2	US-08-456-262-45	Sequence 45, Appl
34	171	6.8	168	2	US-08-456-240-45	Sequence 45, Appl
35	171	6.8	168	2	US-08-455-736-45	Sequence 45, Appl
36	171	6.8	168	4	US-08-971-217-45	Sequence 45, Appl
37	171	6.8	168	4	US-09-350-600-45	Sequence 45, Appl
38	169	6.7	204	1	US-08-614-935-1	Sequence 1, Appl
39	169	6.7	204	3	US-09-130-287-1	Sequence 1, Appl
40	168.5	6.7	912	5	PCT-US95-03747-2	Sequence 2, Appl
41	168	6.6	171	2	US-08-773-368-6	Sequence 6, Appl
42	168	6.6	171	3	US-09-199-887-6	Sequence 6, Appl
43	166	6.6	204	1	US-08-419-414-9	Sequence 9, Appl
44	166	6.6	204	1	US-08-614-935-2	Sequence 2, Appl
45	166	6.6	204	3	US-09-130-287-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-773-368-1
; Sequence 1, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; CORRESPONDENCE ADDRESSES: 6
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PE-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: 1599164
; US-08-773-368-1
Query Match 10.6%; Score 267; DB 2; Length 270;
Best Local Similarity 31.5%; Pred. No. 3.6e-16;
Matches 73; Conservative 28; Mismatches 79; Gaps 8;

QY 13 LLAVALLCTTAAEVPPOLOEQAPMAALNRKESFLLSLHNRLRSWVQPPADMRRL 72
11 LFLPULLLVATT-----GPV-GALTDEKRLMVELHNLRYRAQVSPFASDMLHM 57
QY 73 DMSDSLQAOLQARALCGIPPTPSLASGLMRLQVGMNOLLFAGLASFV-----EV 123
58 RMDEELAAFAKAYV-----RQXRKGHNKRGGRGENLFAITDEGMDVPLA 102
QY 124 VSLMFAEGORYSHAAEGECARNATCTHYTOLWATSSOLGCGRHLC-----SAGOTAIAPV 179
103 MEEWHHEREHYNLAAATCSPOGCGHYTQVWMAKTERIGCGSHFCEKLOGVEETNIELLY 162
QY 180 CAYSPGGMWENVKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPBNP 231
163 CNYEPGN--VKGR--PYOEGTPCSQCPS-----GYHCKNSLCEPIGSP 203
Db

RESULT 2
US-09-199-887-1
; Sequence 1, Application US/09199887
; Patent No. 6071874
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/199,887
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/773,368
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: 1599164
; US-09-199-887-1

Query Match 10.6%; Score 267; DB 3; Length 270;
Best Local Similarity 31.5%; Pred. No. 3.6e-16;
Matches 73; Conservative 28; Mismatches 79; Indels 52; Gaps 8;
QY 13 LLAVALLCTTAAEVPPOLOEQAPMAALNRKESFLLSLHNRLRSWVQPPADMRRL 72
11 LFLPULLLVATT-----GPV-GALTDEKRLMVELHNLRYRAQVSPFASDMLHM 57
Db

QY 73 DMSDSLQAOLQARALCGIPPTPSLASGLMRLQVGMNOLLFAGLASFV-----EV 123
58 RMDEELAAFAKAYV-----RQXRKGHNKRGGRGENLFAITDEGMDVPLA 102
QY 124 VSLMFAEGORYSHAAEGECARNATCTHYTOLWATSSOLGCGRHLC-----SAGOTAIAPV 179
103 MEEWHHEREHYNLAAATCSPOGCGHYTQVWMAKTERIGCGSHFCEKLOGVEETNIELLY 162
QY 180 CAYSPGGMWENVKTIIPYKKGAMCSLCTASVSGCFKAMDHAGLCEVPBNP 231
163 CNYEPGN--VKGR--PYOEGTPCSQCPS-----GYHCKNSLCEPIGSP 203
Db

RESULT 3
US-08-773-368-3
; Sequence 3, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 847722
; US-08-773-368-3

Query Match 9.2%; Score 232; DB 2; Length 219;
Best Local Similarity 32.4%; Pred. No. 3.7e-13;
Matches 58; Conservative 28; Mismatches 61; Indels 32; Gaps 7;

QY 47 ESFL-LLSLHNRLRSWVQPPADMRRLDMSDSLQAOLQARALCGIP-----T 93
11 LFLPULLLVATT-----GPV-GALTDEKRLMVELHNLRYRAQVSPFASDMLHM 80
Db 21 EDFIKDCVRIHNKRREYKPTASDMLYMTWPDALQIAKAWASNOQFSINTRLKPPHKLH 80
QY 94 PSLASGLMRLQVGMNOLLFAGLASFVSVSLMFAEGORYSHAAEGECARNATCTHYTOL 153
81 PNFTS-----LGENWTGSGVPIFSVSSAITNMYDEIODYNFKTRICK--VCGHYTQV 131
Db


```

1 ADDRESS: Incyte Pharmaceuticals, Inc.
2 STREET: 3174 Porter Drive
3 City: Palo Alto
4 STATE: CA
5 COUNTRY: USA
6 ZIP: 94304
7
8 COMPUTER READABLE FORM:
9 MEDIUM TYPE: Diskette
10 COMPUTER: IBM Compatible
11 OPERATING SYSTEM: DOS
12 SOFTWARE: FastSeq for Windows Version 2.0
13
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/09/199,887
16 FILING DATE:
17 CLASSIFICATION:
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: 08/773,368
20 FILING DATE:
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Billings, Lucy RJ
23 REGISTRATION NUMBER: 36,749
24 REFERENCE/DOCKET NUMBER: PF-0186 US
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: 415-855-0555
27 TELEFAX: 415-845-4166
28 TELEX:
29
30 INFORMATION FOR SEQ ID NO: 4:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH: 266 amino acids
33 TYPE: amino acid
34 STRANDEDNESS: single
35 TOPOLOGY: linear
36 IMMEDIATE SOURCE:
37 LIBRARY: GenBank
38 CLONE: 1030053
39
40 US-09-199-887-4
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
98
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,428B
; FILING DATE: 14 No. 564846September 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; OS-08-340-428B-49

```

Query Match	8.4%	Score 212;	DB 1;	Length 1257;
Best Local Similarity	24.8%	Pred. No. 2.4e-10;		
Matches 70; Conservative	34;	Mismatches 134;	Indels 44;	Gaps 10;

QY	177	AFCVAYSGSGNEVNGKTLIPYKKKAMCS--LCTASVSGCFKAMDAHAGLCEVARNPCRM	234
Db	897	ASVSSEGPLRLMDI--PSTLIPVSLGDESDLKVAVESPGLEGFEWEVAASGOEDPTPCEN	955
QY	235	SCQNGG---RLNISTCHCHCPRGYTRYCOV-----RCSLQCVAGREFREECS--CYCD	283
Db	956	NPCLTGTCRTNGTMYGSCSDQYAGENCEIDIDDLCLCS--PCENGGTCIDEVNGTICCL	1011
QY	284	ICYGGAOCATKYNHFFPHDCL---RLDGCDFWVSSEADLYYRAAMKQCRKGGVLAQLKSQ	340
Db	1015	PSYGGNLECKDT---EGGCDRGMHKQGHYRYFAHRAAMEDARDORRRRGHILTSVSP	1070
QY	341	KVODLFAEYGLRLETTNEVTDSDFFETRNWVIGLTYKAKDSFRVATGEHQAFTSFAFGCP	400
Db	1071	EEHKFI-----NSFGHENSIGLINDRIVERDFQWTDNTMGLOIENWMREKQP	1119
QY	401	DNHGLVWLSAAMFGNCEVLAQSAATNNWQDQCKTRRRTYICQ	442
Db	1116	DNF-----FAGGEDCVYVAHENGRRMDVPCPNYILPYCK	1150

```

RESULT 8
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI T.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO:2:
; LENGTH: 2409
5180808-2

Query Match          7.83; Score 196; DB 6; Length 2409;
Best Local Similarity 26.83; Pred. No. 1.5e-06;
Matches 63; Conservative 27; Mismatches 83; Indels 62; Gaps 12;

y 230 NPCRMSQNGRLNISTCH-----CHCPGGYGRYQGV-----RSLQCVHG- 271

```

Db 2109 NRC-----LNGTCTPTETSYVCTGVPYSGDQCELDPECHSNPCRNATCVDGF 2159
Qy 272 -RREBECSCVDIGYGACATKVFPHFTCDL---RIGDCEFMWSEADTYRARKC 327
Db 2160 NFR-----CLCLPSYVVALCEQDT---ETCOYGMHKPGQCKYKAHRTTWAAREC 2210
Qy 328 QKRGVLAOKSOKVODILAFYIGLETTEWVDSDEFTRNFWIGLYTKAKDSFRWATG 387
Db 2211 RIQGAHLTITILSHEDQ---MEVNRV-----GHDXQ---WIGLNDKMEHFHRTWDG 2255
Qy 388 EHQAFSFAFGPDNIGLWLSAAMGFNCVELQASAFWMDORCKTRNRITYQ 442
Db 2256 STLOYLENRPNDPS---FFSAG---EDCVIITHENGQMDVPCNVHLTYTCK 2303

RESULT 9

US-08-840-062-5
Sequence 5, Application US/08840062
Patent No. 6117977
GENERAL INFORMATION:
APPLICANT: LASKY, LAURENCE A.
APPLICANT: WU, KAI
TITLE OF INVENTION: TYPE C LECTINS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/840,062
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1019R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1455 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-840-062-5

Query Match 7.5%; Score 189.5; DB 3; Length 1455;
Best Local Similarity 23.3%; Pred. No. 3e-08;
Matches 84; Conservative 45; Mismatches 105; Indels 127; Gaps 17;

Qy 135 SHA-----AGECARNATCH--YTOLVWATSSQLGCGRHLC---AGQTAIE-----AFV 179
Db 491 SHAHVPEGADKCRGKRWKNGHGYCYLIGSTLSTFTDANHCTNEKAYLITVEEDREYEQAF 550
Qy 180 CA-----YSPGCMWEYNGK---TI-----IPYKKGAMCSLCTASVSGC 214
Db 551 TSLVGLRPKRYWTGSDONKGTFTWYDEQYQFTHWANADMGKRAK---CVAMKTV 606
Qy 215 EKAMHAGGLCEVPRNCPMCSQNHGRNLNITCCHCPCPGYGRYQVRCSLQCVHGRFR 274
Db 607 -----AGGLMDV-----LSC----- 616
Qy 275 EEECSQVCDIGCGAGCAIKVHFPHFTCDLRIDG-----CFMV-----SSEADTY 321

Db 617 EEKAFVCK---HMAEGYRPEPTPTPEPKCPENNCTTSKTSMPCKYAKGHEKKTWF 673
Qy 322 RARMKOQRKGVLAOKSOKVODILAFYIGLETTEWVDSDEFTRNFWIGLYTKAKDS 381
Db 674 ESRDFKALIGGLASTKSDQEOV---IWRL-----ITSSGYHELFWGLTYGSPSEG 724
Qy 382 FRWANGHQAFSFAFGQPDNIGLWLSAAMGFNCVELQASAFWMDORCKTRNRITYC 441
Db 725 FTWDSGSPYSYNWAVGERNNY-----QNYEYCGELKGDGMSMNDLNCHLNWIC 776
Qy 442 Q 442
Db 777 Q 777

RESULT 10

US-08-614-935-28
Sequence 28, Application US/08614935
Patent No. 5804201
GENERAL INFORMATION:
APPLICANT: King, Te P.
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,935
FILING DATE: 11-MAR-1996
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-156
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 151 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-614-935-28

Query Match 7.3%; Score 185; DB 1; Length 151;
Best Local Similarity 30.5%; Pred. No. 3.7e-09;
Matches 47; Conservative 28; Mismatches 67; Indels 12; Gaps 5;

Qy 51 LLSLHNRRLRSWQPPAAMRRRLDWSLSLAQLAARALAGI-----PTPLASGLMRTLOY 106
Db 6 IVNKNELRKAVSPSPASMLKMWREBVTYTNQRRANKCTTLQHSDEDEKTS-----TIC 60
Qy 107 GWNQQLPAGLASVEYVSLWFAEGORYSHAAGECARNATCTHYTQLVWATSSQLCGGRH 166

```

Db      61  GENL-YMSDPTSMSSAIAQSWDEILDFEYGVGPKSPBNNAVGHYTLQWVSTYGVGCGIA 119
QY      167  LCSAGQTAIEAFVCAVSPGKNWEVNGKTIIPYK 200
Db      120  YCPNODSLKYYVVCQYCPAGN-NMNRKN-TPYQQ 151

RESULT 11
US-09-130-287-28
; Sequence 28, Application US/09130287
; Patent No. 6106844
; GENERAL INFORMATION:
; APPLICANT: King, Te P.
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
; TITLE OF INVENTION: ANTIGEN 5
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/130,287
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/614,935
; FILING DATE: 11-MAR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-156 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; ..
US-09-130-287-28

Query Match          7.3%; Score 185; DB 3; Length 151;
Best Local Similarity 30.5%; Pred. No. 3.7e-09;
Matches 47; Conservative 28; Mismatches 67; Indels 12; Gaps 5

QY      51  LLSLHNRLRSWVQPPAADRLRLDMSDSLQAQARAALCGI----PPPSIASGLMRTIQQV 106
Db      6  IVNKHNELRKAVSPASNMLKMEWSREVTNTAQRMANCKTLQHSDEPDRKS-----TRC 60
QY      107  GWNQQLPAGLASPEVEVNSLMPAEGCRYSIAGECARNACTHTLTQLVMTATSSDLGCGRH 166
Db      61  GENL-YMSDPTSMSSAIAQSWDEILDFEYGVGPKSPBNNAVGHYTLQWVSTYGVGCGIA 119
QY      167  LCSAGQTAIEAFVCAVSPGKNWEVNGKTIIPYK 200
Db      120  YCPNODSLKYYVVCQYCPAGN-NMNRKN-TPYQQ 151

```

```

RESULT 12
US-07-857-224B-97
: Sequence 97, Application US/07857224B
: Patent No. 5958784
:
: GENERAL INFORMATION:
: APPLICANT: Benner, Steven A.
: TITLE OF INVENTION: Predicting Folded Structures of Proteins
: NUMBER OF SEQUENCES: 114
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Steven A. Benner
: STREET: Hadlaubstrasse 151
: CITY: Zurich
: STATE: none
: COUNTRY: Switzerland
:
: ZIP: (note: this is an international post code) CH-8092
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage
: COMPUTER: Apple Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/857,224B
: FILING DATE: 03/25/92
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA: none
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (International) 41 1 632 2830
: TELEFAX: (International) 41 1 262 2437
:
: TELEX: none
:
: INFORMATION FOR SEQ ID NO: 97:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 135
: TYPE: amino acid
: TOPOLOGY: linear
:
: MOLECULE TYPE:
:
: DESCRIPTION: protein
:
: ORIGINAL SOURCE:
:
: ORGANISM: tobacco
:
: FEATURE: Pathogenesis related protein, Table 16 Row 1
:
: PUBLICATION INFORMATION:
:
: AUTHORS:
:
: AUTHORS: Cutt, J. R.
: AUTHORS: Dixon, D. C.
: AUTHORS: Carr, J. P.
: AUTHORS: Klessig, D. F.
:
: TITLE: Isolation and nucleotide sequence of cDNA clones for the
: TITLE: pathogenesis related proteins of Nicotiana glauca induced by TMV
:
: JOURNAL: Nucleic Acids Research
:
: VOLUME: 16
:
: PAGES: 9861-
:
: DATE: 1988
:
US-07-857-224B-97

```

	Query Match	Similarity	7.2%	Score 182.5	DB 2:	Length 135;
	Best Local	Similarity	33.3%	Pred. No. 5.3e-09;		
	Matches	47;	Conservative	17;	Mismatches	54;
					Indels	23;
					Gaps	5
Oy	52	LSLHNRKRSWVPPADMRRLDMSDLAQLAO-----ARAALGIGIPTPLSLASGLMRTLQVG	107			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	8	LAVIHNDARAQV-----GVGPMSMDANILSRAGNYANSRAGDCNL-----IHSAGEMLANG	58			
Oy	108	NNMQLLPGLASFVEYVSLSIFPAEGORYSIAGCECARNAATCTHYTLQWLMTSSQLGGCRHL	167			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	59	-----GGDFTGRAAYOLWTSERPSYNATNCQCGKKRCRHYYQTQVMENSVVLGGCGRAR	111			
Oy	168	CSAGOTALEAFVCASPCCGNW	188			
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	112	CNNGWFIS--CNYDPVCGNW	129			

RESULT 13

US-07-857-224B-98

Sequence 98, Application US/07857224B

Patent No. 5958784

GENERAL INFORMATION:

APPLICANT: Benner, Steven A.

TITLE OF INVENTION: Predicting Folded Structures of Proteins

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Steven A. Benner

STREET: Hadlaubstrasse 151

CITY: Zurich

STATE: none

COUNTRY: Switzerland

ZIP: (note: this is an international post code) CH-8092

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 7.0

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/857, 224B

FILING DATE: 03/25/92

CLASSIFICATION: 436

PRIOR APPLICATION DATA: none

TELECOMMUNICATION INFORMATION:

TELEPHONE: (International) 41 1 632 2830

TELEFAX: (International) 41 1 262 2437

TELEX: none

INFORMATION FOR SEQ ID NO: 98:

SEQUENCE CHARACTERISTICS:

LENGTH: 135

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: protein

ORIGINAL SOURCE:

ORGANISM: Lycopersicon esculentum

FEATURE: Pathogenesis related protein; Table 16 Row 2

PUBLICATION INFORMATION:

AUTHORS:

AUTHORS: Cutt, J. R.

AUTHORS: Dixon, D. C.

AUTHORS: Carr, J. P.

AUTHORS: Klessig, D. F.

TITLE: Isolation and nucleotide sequence of cDNA clones for the

TITLE: pathogenesis related proteins of Nicotiana tabacum induced by TMV

TITLE: infection.

JOURNAL: Nucleic Acids Research

VOLUME: 16

PAGES: 9861

DATE: 1988

US-07-857-224B-98

Query Match

Best Local Similarity 7.2%; Score 182.5; DB 2; Length 135;

Matches 47; Conservative 17; Mismatches 54; Indels 23; Gaps 5;

52 LSHNRLRSWVPPADMRRLMSDLSIAQLAQ---ARALCGIPTPSLASGLWRTLOVG 107

8 LAVHNDARQV-----GVGPMWDANLASRAQYANRAGDCNL---IHSGAGENLAKG 58

QY 108 WNMQLPAGIASFEVVSILMFAGORYSHAAGECARNATCTHYTOLVWATSSQLGCGRHL 167

Db 59 -----GGDFTGRAVQVLWVSERPTYNATNOCVGGKCRHRYQVWVRNSVRLGCGRAR 111

QY 168 CSAGCTAIEAFVCAVSPGGM 188

Db 112 CNGGMWFTS---CNYDPVGNW 129

RESULT 14

US-07-857-224B-98

Sequence 99, Application US/07857224B

Patent No. 5958784

GENERAL INFORMATION:

APPLICANT: Benner, Steven A.

TITLE OF INVENTION: Predicting Folded Structures of Proteins

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Steven A. Benner

STREET: Hadlaubstrasse 151

CITY: Zurich

STATE: none

COUNTRY: Switzerland

ZIP: (note: this is an international post code) CH-8092

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh 7.0

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/857, 224B

FILING DATE: 03/25/92

CLASSIFICATION: 436

PRIOR APPLICATION DATA: none

TELECOMMUNICATION INFORMATION:

TELEPHONE: (International) 41 1 632 2830

TELEFAX: (International) 41 1 262 2437

TELEX: none

INFORMATION FOR SEQ ID NO: 99:

SEQUENCE CHARACTERISTICS:

LENGTH: 135

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

DESCRIPTION: protein

ORIGINAL SOURCE:

ORGANISM: Lycopersicon esculentum

FEATURE: Pathogenesis related protein; Table 16 Row 3

PUBLICATION INFORMATION:

AUTHORS:

AUTHORS: Cutt, J. R.

AUTHORS: Dixon, D. C.

AUTHORS: Carr, J. P.

AUTHORS: Klessig, D. F.

TITLE: Isolation and nucleotide sequence of cDNA clones for the

TITLE: pathogenesis related proteins of Nicotiana tabacum induced by TMV

TITLE: infection.

JOURNAL: Nucleic Acids Research

VOLUME: 16

PAGES: 9861

DATE: 1988

US-07-857-224B-98

Query Match

Best Local Similarity 7.1%; Score 180.5; DB 2; Length 135;

Matches 47; Conservative 17; Mismatches 54; Indels 23; Gaps 5;

52 LSHNRLRSWVPPADMRRLMSDLSIAQLAQ---ARALCGIPTPSLASGLWRTLOVG 107

8 LAVHNDARQV-----GVGPMWDANLASRAQYANRAGDCNL---IHSGAGENLAKG 58

QY 108 WNMQLPAGIASFEVVSILMFAGORYSHAAGECARNATCTHYTOLVWATSSQLGCGRHL 167

Db 59 -----GGDFTGRAVQVLWVSERPTYNATNOCVGGKCRHRYQVWVRNSVRLGCGRAR 111

QY 168 CSAGCTAIEAFVCAVSPGGM 188

Db 112 CNGGMWFTS---CNYDPVGNW 129

RESULT 15

US-08-419-414-10

Sequence 10, Application US/08419414

Patent No. 5753787

GENERAL INFORMATION:

APPLICANT: Hawdon, John M.
APPLICANT: Hotez, Peter J.
APPLICANT: Jones, Brian F.
TITLE OF INVENTION: Hookworm Vaccine
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/419,414
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: YU113
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8795
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Polistes annularis
US-08-419-414-10

Query Match 7.1%; Score 180.5; DB 1; Length 205;
Best Local Similarity 30.8%; Pred. No. 1.4e-08;
Matches 53; Conservative 24; Mismatches 70; Indels 25; Gaps 7;

QY 46 KESFLLSLNRLRSWV-----OPPADMRRLDMSDSLALQARALCGIPT 93
DB 41 EEKKLIYSEHNRRPQKVAQGLETRGNRPQRPASDMNDLVWDELALHIAQVWASQCF-- 98
QY 94 PSLASGLMRT--LQVGMNQLL-PAGIASFEVYVSLMFAEGORYSHAAGECARN-ATCT 148
DB 99 -LVHDKCRNTAKPYVGNINAYAGSINLPDVVSLIKLWENEVKDFNYNTGITKQNFAGIG 156
QY 149 HTYQVWATSSQAGCGRHLCSAGQTAIFACVSPGGMWEVNGKTIIPYK 200
DB 157 HTYQVWATSSQAGCGRHLCSAGQTAIFACVSPGGMWEVNGKTIIPYK 204

Search completed: December 29, 2002, 02:27:28
Job time : 39 secs